METHODS OF MEASURING MICROSCOPIC TISSUE DAMAGE IN CANCELLOUS BONE: SAMPLING AND STATISTICAL POWER

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Katherine Marie Ehlert

August 2013

ABSTRACT

Microscopic tissue damage can occur in bone as a result of an isolated overload leading to reduces bone strength under subsequent loads. In addition, microscopic tissue damage is believed to stimulate bone resorption and bone loss. Microscopic tissue damage in bone is most commonly measured in two-dimensional sections using stereology techniques. Stereology techniques are accurate but can have limited precision. Previous studies have presented guidelines for adequate sampling of naturally occurring microcracks in cortical bone, but such guidelines have not been presented for cancellous bone and have not addressed other forms of microscopic tissue damage (i.e. diffuse damage). Here a statistical model is presented that can be used to design studies in which microscopic tissue damage in cancellous bone is a key study outcome.

BIOGRAPHICAL SKETCH

Katherine Ehlert is the third child of George and Georganne Ehlert. Katherine was born and raised in a small suburb of Detroit, MI with her two older brothers. After graduating from South Lyon High School, Katherine attended Case Western Reserve University in Cleveland, OH where she majored in Mechanical Engineering and focused her studies on materials and biomechanics. After Case Western, Katherine attended Cornell University to pursue a Master's Degree in Mechanical Engineering under the advisement of Dr. Christopher Hernandez. Her main focus while attending Cornell has been on quantifying and understanding the mechanisms of microscopic damage accumulated during mechanical testing of human cancellous bone. At the time of publication, Katherine is a consulting engineer at Exponent, Inc. in their Biomedical Engineering Practice. Her focus is mechanical characterization of cadaveric and animal tissue including bone, muscle, skin, and bioengineered materials.

For my parents, my brothers, my extended family, and my friends. Your unwavering love and support throughout the years is the reason this research exists today.

ACKNOWLEDGMENTS

This thesis was made possible by Grant Number AR057362 from NIAMS/NIH. Its contents are solely the responsibility of the author and do not necessarily represent the official views of the NIAMS or NIH. The author would also like to thank Ralph O'Brien for his contributions to the statistical modeling software, the many great conversations, and pieces of advice. The author would also like to thank Floor Lambers, and Stephanie Dux for access to their experimental data and analysis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1.0 INTRODUCTION

Microscopic tissue damage can form in bone tissue during normal activities or isolated overloading events. Microscopic tissue damage has the potential to reduce bone strength under subsequent loading [1] and has been shown to stimulate bone resorption in vivo [2]. While a great majority of studies of microscopic tissue damage have concentrated on naturally occurring microdamage in otherwise normal bone tissue, examination of microscopic tissue damage generated during biomechanical testing in vitro can provide information regarding failure processes in bone tissue. Induced microscopic tissue damage can provide insight of how failure processes are influenced by aspects of bone that can be altered by drug treatments. Possible parameters that can be altered with drug treatments include local bone morphology (trabecular microarchitecture, resorption cavities, etc.) or local variation in tissue properties (osteons, cement lines, tissue degree of mineralization, collagen cross-linking, etc.).

The current standard for measuring microscopic tissue damage in bone is bulk stain followed by sectioning (50-200 μm thick) and quantification using stereology techniques. Microscopic tissue damage is characterized in two forms: 1) microcracks (80 - 150 μm in length) measured as the crack density (Cr.Dn, number of cracks per unit bone cross-sectional area); and 2) diffuse damage, a region of diffuse staining measured through point counting and expressed as a proportion of the total amount of bone (DV/BV). The term “microdamage” is used to describe both microcracks and diffuse damage. Recently, three-dimensional images of bone and microdamage have

been developed using micro-computed tomography (with radio-opaque stains) as well as with serial milling (using fluorochrome stains) [3-6].

A challenge to understanding the contribution of microdamage to mechanical failure of cancellous bone is that measures of microdamage are highly variable [7]. Studies of microdamage generated in vitro in cancellous bone specimens often show coefficient of variations (SD/Mean) in excess of 0.60 (Table 1). The large variability in microdamage limits statistical power. For example, a measurement with a coefficient of variation of 0.60 would require 143 specimens per group to detect a 30% difference between two groups using a t-test ($\alpha = 0.05$ and statistical power = 0.90). As 143 specimens per group is not feasible in biomechanical studies (few biomechanical studies use more than 30 specimens), methods of improving statistical power by altering the measurement approach (number of sections, cross-sectional area examined) or study design (varying the number of donors, number of specimens per donor, and the amount of applied mechanical load) are needed to better understand the relationship between microdamage, mechanical performance, and mechanical failure in cancellous bone. Improvements in measurement approach reduce the contribution of measurement error to variance while alterations in study design can make it possible to obtain more insight from an experiment for a given number of specimens. It is not known exactly how measurement technique influences measurement error in cancellous bone nor is it known how study design influences the ability to detect the effects of an experimental stimulus.

Compilation of Studies that Report Microscopic Tissue Damage Measured Using Two-Dimensional Sections

Table 1. Microscopic tissue damage generated by a controlled mechanical load is shown. Each test group in the experiment is reported in separate line. Most studies showed coefficients of variation (CV = SD/Mean) greater than 0.50. Values that were not expressly reported by the authors were left as NR.*11 different groups were tested in this study, the median value is reported. ^22 different groups were tested in this study, the median value is reported.

Author (Year)	**Technique**	**Sections Analyzed**	**Area Analyzed**	**Damage Fraction ± SD**	CV
O'Neal (2010) [16]	Damage Incidents	6	NR	0.74 ± 0.14	0.19
				0.43 ± 0.24	0.56
Dux (2010) [17]	Stereological Point Counting on Slides	2	42 mm^2	0.16 ± 0.20	1.26
				0.26 ± 0.31	1.20
Nagaraja (2007) [18]	Automatic Segmentation	4	32 mm^2	0.0023 ± 0.0012	0.53
				0.0018 ± 0.0012	0.68
				0.011 ± 0.0051	0.46
				0.010 ± 0.0022	0.21
Nagaraja (2005) [19]	Stereological Point Counting on Images	4	NR	0.011 ± 0.0073	0.61
				0.010 ± 0.0031	0.31
Wang (2006) [20]	Tracing of Image	2	NR	0.0046 ± 0.0017	0.38
				0.0044 ± 0.0014	0.31
				0.0046 ± 0.0017	0.37
Wang (2005) [21]	Tracing of Image	2-4	84 mm^2	0.0045 ± 0.0014	0.31
Moore (2003) [22]	Tracing of Image	NR	118 ± 40 mm^2	0.0040 ± 0.0033	0.82*
Moore (2002) [23]	Tracing of Image	3-5	301 ± 63 mm^2	0.018 ± 0.13	0.72^

Compilation of Studies that Report Microscopic Tissue Damage Measured Using Three-Dimensional Analysis Techniques

Table 2: Microscopic tissue damage reported in previous studies in which a controlled mechanical load was applied in vitro and three-dimensional measures of microdamage were used is shown. Each test group in the experiment is reported in separate line. Most studies showed coefficients of variation (CV = SD/Mean) greater than 0.50. Values that were not expressly reported by the authors were left as NR.

Author (Year)	3D Technique	Voxel Size	Volume Analyzed	Damage Fraction ± SD	CV
Wang (2007) [4]	$BaSO_4$ Stain with µCT	10 x 10 x 10 µm	125 mm^3	0.045 ± 0.029	0.65
				0.012 ± 0.0081	0.66
Tang (2010) [10]	Lead Uranyl-Acetate with µCT	10 x 10 x 10 µm	64 mm^3	0.15 ± 0.022	0.14
				0.14 ± 0.018	0.13
Karim (2011) [11]	Lead Uranyl-Acetate with µCT	17.5 x 17.5 x 17.5 µm	8 mm^3	0.059 ± 0.061	1.03
				0.026 ± 0.025	0.99
				0.0026 ± 0.0038	1.46
Bigley (2008) [5]	Serial Block-Face Imaging	9 x 9 x 5 μm^3	NR	0.047 ± 0.021	0.45
Slyfield (2012) [6]	Serial Block-Face Imaging	2.8 x 2.8 x 5 μm^3	190 mm^3	0.77 ± 0.33	0.43

Martin and colleagues presented a statistical model to explain how the number of sections taken from each specimen and cross-sectional area examined in each section influences measurement variance in cortical bone [8]. The study provided key information regarding how to quantify naturally occurring microdamage in cortical bone but did not address quantification of diffuse damage or quantification in cancellous bone and did not explicitly report statistical power. Here, the work by Martin and colleagues has been generalized to examine diffuse damage measures, and allow for simulation of all densities of bone (both cancellous as well as cortical bone).

The overall goal of this line of work is to understand the relationship between microscopic tissue damage and cancellous bone biomechanical performance. Specifically in this study the following questions were asked: (1) how does the cross-sectional area examined in each section, the number of sections examined, and the amount of microdamage in the specimen affect measurement error; (2) how do changes in study design (number of donors, number of specimens per donor, distribution of specimens in a study) affect the ability to describe the relationship between the amount of microdamage and applied mechanical load (or other predictor)? The current work differs from prior work in that it is applied to cancellous bone, it considers measures of diffuse damage, and addresses methods of allocating specimens among experimental groups that are based on the magnitude of a stimulus (in this case applied mechanical load).

2.0 THEORY

Statistical simulation is a common approach used to design studies or calculate statistical power when closed form solutions are not available. To perform such an analysis one first uses pilot data to describe the statistical distribution for the population of interest and then simulates an experiment by randomly selecting quantitative data from that distribution. By repeating the simulation thousands of times a distribution of possible results for an experiment is achieved. By modifying parameters such as number of sections analyzed or number of specimens, it can determined how these parameters influence the distribution of possible results and characterize the likelihood that a proposed experimental design will detect differences between groups.

2.1 Model 1 – Measurement Error Analysis

In most studies microdamage in bone is measured in a finite number of two-dimensional sections collected from each specimen. Measurement variability can occur because the entire sample is not measured and only specific regions of each section are examined. To determine how measurement technique (cross-sectional area per section, number of sections) influences measurement error, a statistical model written for use with R (www.r-project.org, see Appendix 2) was implemented. The model assumes a population distribution of microdamage across individuals defined by the median and the intra-95% range (RS_{95}, the ratio of the 97.5^{th} percentile to that of the 2.5^{th} percentile). The median and RS_{95} are analogous to the mean and standard deviation but characterize non-normal distributions.

Following the work by Martin and colleagues [8], microdamage within a specimen cross-section was simulated using a Poisson distribution:

$$P(k) = e^{-\mu} \frac{\mu^k}{k!} \quad (1)$$

where P(k) is the probability that a cross-section contains k instances of microdamage (cracks or a patch of diffuse damage) and μ is the mean number of instances of microdamage per cross-section. The model assumes measurements are made on each section accurately (i.e. there is no observer error in measurement), that each observed instance of microscopic tissue damage is independent, and that microdamage is equally likely to occur throughout the specimen. The statistical model considers measurement of crack density (Cr.Dn, microcracks per unit bone area) separately from that of diffuse damage to account for differences in variance due to the fact that crack density is a ratio of a discrete variable (the number of cracks within a specimen) to a continuous variable (the bone area) while diffuse damage is determined as the ratio of two continuous variables (DV/BV).

The model simulates the measurement of microcracks by first randomly selecting a crack density value from the population distribution to represent the true mean crack density in one specimen (True.Dx, Figure 1). The true mean crack density of that specimen is then used to create a Poisson distribution of the number of cracks observed in a section (Equation 1, with a mean value μ = True.Dx). The average number of microcracks per unit area observed within one cross-section is calculated and stored in the model. The simulation is then repeated for each cross-section examined within the specimen and the results for each cross-section are averaged and

reported as the measured crack density (Meas.Dx) for the specimen. Measurement error is calculated as the ratio of the measured crack density in the specimen and the true crack density ($\text{Error} = \frac{\text{Meas.Dx}}{\text{True.Dx}}$, a value of 1.0 signifies there is no measurement error). The whole process is repeated 5000 times (simulating analysis of 5000 specimens) and the 2.5% and 97.5% quantiles of the measurement error are reported (Figure 1). A similar approach is used to examine diffuse damage, however, a Poisson distribution models discrete events, thus an extra parameter is necessary for the diffuse damage simulation ($M.RS_{95}$). The $M.RS_{95}$ parameter adjusts the diffuse damage measure such that it can be represented by a Poisson distribution. Measurement techniques (number of sections, cross-sectional area per section, etc.) that result in the smallest intra-95% range of measurement error are most accurate.

Three-dimensional measurements of microdamage (radio-opaque stains with micro-CT or fluorescent stains with serial milling) are simulated by considering each cross-section in the three-dimensional image to be an individual section. As an example we consider a cylinder of cancellous bone that is 8mm in diameter and 5mm in length (a typical specimen for mechanical testing in our laboratory) which is imaged in three-dimensions using 1000 cross-sections (5 μm in thickness), each with a cross-sectional area of 50.27 mm^2.

Schematic Representing the Measurement Error Analysis Software for Measuring Crack Density

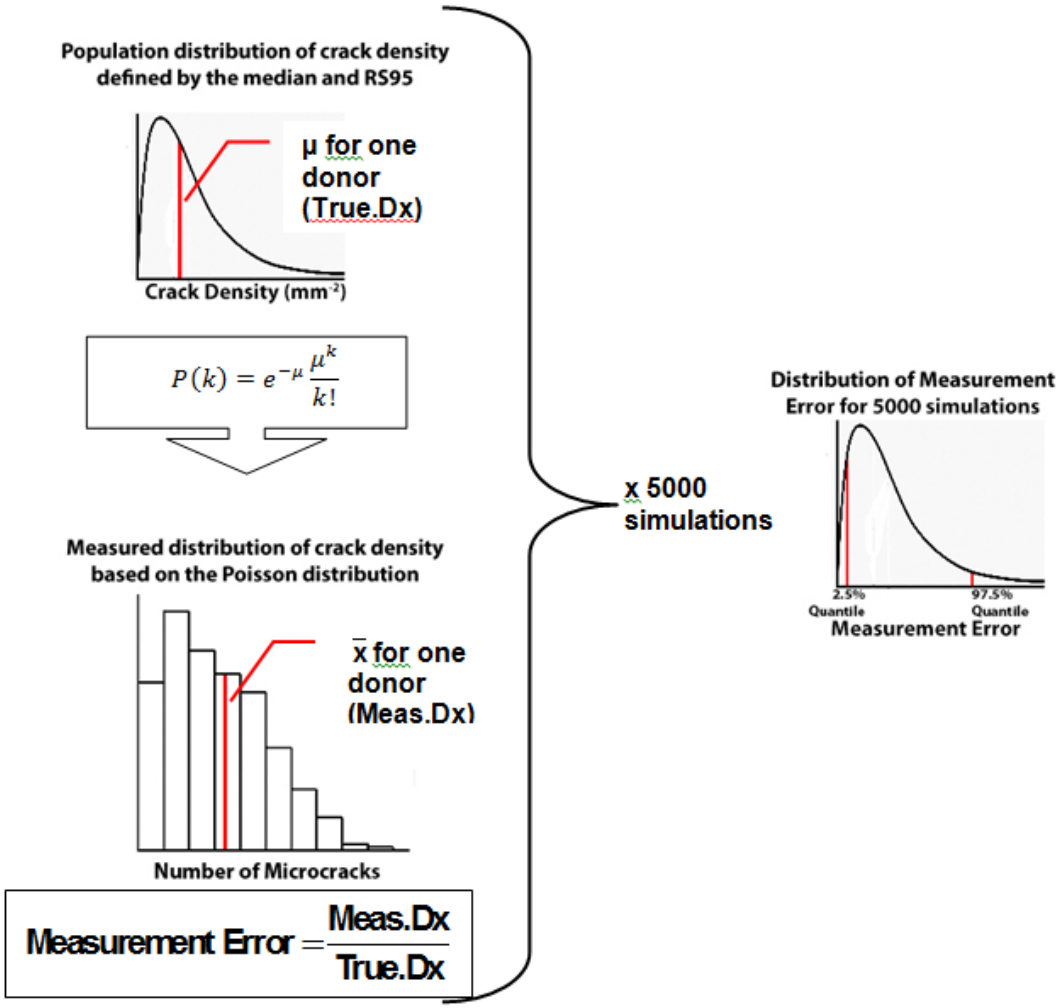

Figure 1: The method of simulating microdamage is illustrated. A population distribution of specimens based on previous experiments is assumed. From that population distribution, a mean crack density value (True.Dx) is randomly selected and used to define the Poisson distribution which simulates each microcrack observed in a cross-section. The number of microcracks for each cross-section is averaged for the number of sections analyzed. The average estimated crack density (Meas.Dx) is compared to the real crack density (True.Dx) to determine measurement error (Meas.Dx/True.Dx). The simulation is repeated 5000 times and the intra-95% range of measurement error is reported.

2.2 Model 2 – Number and Distribution of Study Groups

Measurement of microdamage is often required to compare two or more experimental groups and/or identify relationships between a stimulus and the resulting amount of microdamage. A common experimental design is to have two or more experimental groups, each experiencing a different magnitude of stimulus. To identify the most attractive distribution of number of donors, number of specimens per donor, and magnitude of stimulus a statistical model written for use with R (www.r-project.org, see Appendix 2) was implemented. In the current analysis applied mechanical strain was the only stimulus considered. It is important to note that by modifying specific parameters, however, the same software can be used for any independent parameter (donor age, microarchitecture, etc.). When there are only two experimental groups, the analysis is equivalent to a t-test (Figure 2) [9].

Schematic of a T-Test Implemented as an Ordinary Linear Regression Model

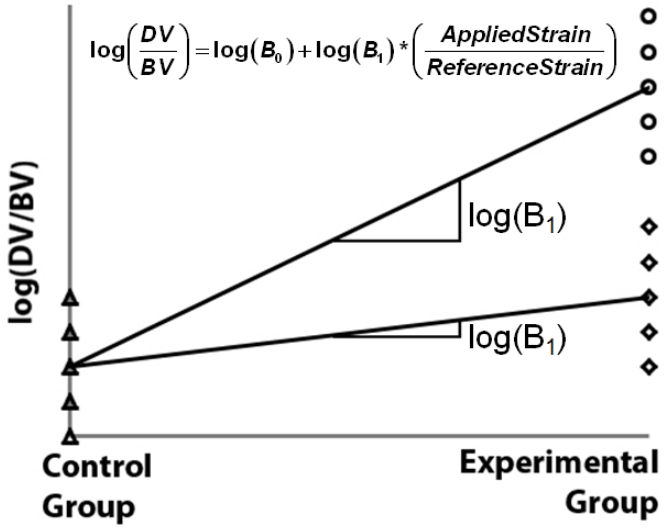

Figure 2. A graphical representation of a t-test implemented as an ordinary linear model is shown. Two tests are represented, one with a 60% increase in the Treatment Group (circles) and one with a 10% increase in the Treatment Group (diamonds). The y-intercept ($\log(B_0)$) is the mean amount of microdamage in the control group and the

slope of each line ($\log(B_1)$) is the difference between means. A steeper slope indicates a larger difference between groups.

Over a range of applied mechanical strains (0% applied strain – 2.25% applied strain), the relationship between microdamage and applied strain is exponential [3]. The following equation was then used to express the exponential relationship between the applied strain and the amount of microdamage present in cancellous bone:

$$\left(\frac{DV}{BV}\right) = B_0 * B_1^{\left(\frac{\text{AppliedStrain}}{\text{Ref.Strain}}\right)} \tag{2}$$

where B_0 is the mean amount of damage in the unloaded control group (y-intercept of the curve fit) and B_1 is the increase in damage volume for a particular applied strain (the slope of the relationship on a log-log axis). The reference value (Ref.Strain) is an arbitrary scaling factor and was chosen to be within our applied strain range (2% applied strain). By applying a logarithmic transformation we get the following linear relationship:

$$\log\left(\frac{DV}{BV}\right) = \log\left(B_0\right) + \log\left(B_1\right) * \left(\frac{\text{AppliedStrain}}{\text{Ref.Strain}}\right) \tag{3}$$

For the simulations, the statistical power for a study design was determined by assuming a population distribution and simulating thousands of individual experiments using the following steps: First, population distributions for the parameters B_0 and B_1 are assumed (using preliminary data and/or previous experience) and expressed in terms of median and RS_{95}. Measurements in one individual are simulated by selecting values for B_0 and B_1 from the population distributions at random. The selected B_0 and B_1 values define the true relationship between applied strain and microdamage in one individual (True.B_0, True.B_1, respectively). Once B_0

and B_1 are determined, Equation 2 is used to express the amount of microdamage (True.Dx) expected for an applied mechanical strain. The True.Dx parameter represents the damage volume if the donor sample could be perfectly analyzed. Since, as we have seen from Model 1, no measurement is perfect, noise is added to simulate an actual experiment (see Appendix 1 for details). The predicted damage volume (Meas.Dx) is determined for each experimental group and used to determine the measured relationship between applied mechanical strain and microdamage in the simulation (Meas.Rel). The simulation is repeated 5000 times and a distribution for B_0 and B_1 are determined (Meas.B_0 and Meas.B_1, respectively) (Figure 3).

Schematic Representing the Analysis Software for Determining the Number and Distribution of Experimental Groups when Measuring the Relationship Betweenn Applied Strain and the Accumulation of Microdamage

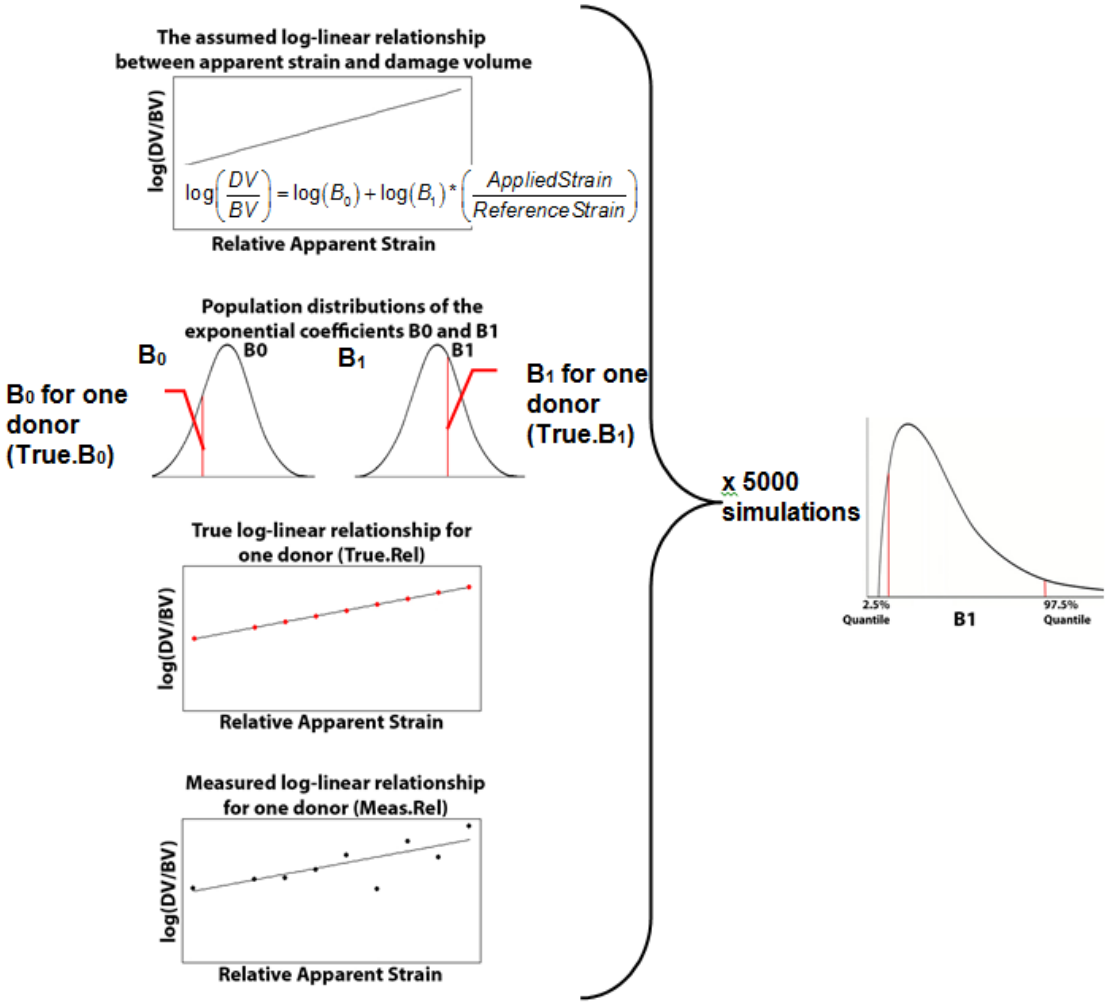

Figure 3. A depiction of the simulation process for the regression model experiments is shown. The model begins with an assumed relationship between the stimulus (mechanical loading) and response (microdamage accumulation). Population distributions are then defined for each coefficient in the relationship. The coefficient values are then randomly selected from the defined population distributions (True.B_0 and True.B_1, respectively) to create the true exponential relationship of a donor (True.Rel). Based on the measurement error values from model 1 and previous experiments, noise is added to the true value at each experimental group. An exponential fit for the data is calculated (Meas.Rel) to estimate the B_0 and B_1 coefficients (Meas.B_0 and Meas.B_1, respectively). The process is then repeated 5000 times to represent a full population of experiments and a distribution of the estimated B_1 is reported.

A total of 5000 experiments measuring microdamage are simulated. The 95% confidence interval of the parameter B_1 is determined. The parameter B_1 expresses the slope of the model in Equation 2 and is equal to one if there is no difference between two groups or no relationship between the independent and dependent variables in the case of more than one experimental group. Statistical power expresses the probability of detecting a difference in B_1 from the null value for a given value of α (in our case $\alpha = 0.05$). In the examples described below, a value of $B_1 = 5$ is used as the null value rather than the more traditionally used value of $B_1 = 1$. Values of B_1 exceeding 5 indicate not only that the slope of the curve significantly different from one (i.e. a trend exists), but that there is a convincing relationship between applied mechanical strain and the accumulation of microdamage. Statistical power was calculated using the hypothesis that $B_1 > 5$ with $\alpha = 0.05$. Statistical power identified with this more conservative null hypothesis will be referred to as 'essential statistical power' through the remaining portion of the thesis.

3.0 RESULTS

3.1 Model 1 – Measurement Error Analysis

3.1.1 Cross-sectional Area

The cross-sectional area examined in each two-dimensional section varies considerably among laboratories and studies (range 30 - 300 mm^2) and there are no clear recommendations for such analysis examining microdamage in cancellous bone (Table 1). With regard to crack density, increases in cross-sectional area per section reduced measurement error (as indicated by the reductions in size of the 95% confidence interval, Figure 4). However, increasing the cross-sectional area per section beyond $40mm^2$ had only a minor effect on measurement error. The measurement error of diffuse damage was not affected by alterations in cross-sectional area per section.

The Effect of Cross-sectional Area and Bone Volume Fraction on Measurement Error for Crack Density

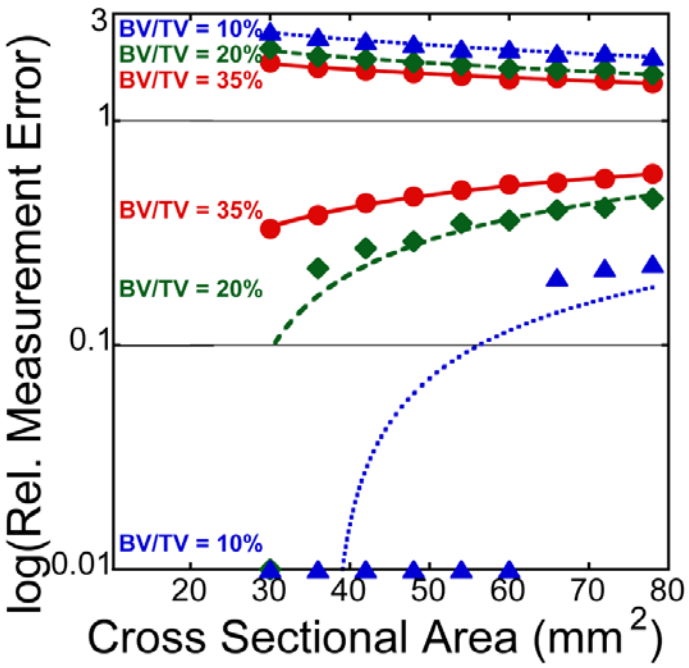

Figure 4. The intra - 95% range of measurement error for crack density is illustrated as the distance between points in the vertical direction. Increasing the cross-sectional area examined in each specimen reduces the range, however, when the cross-sectional area analyzed is increased beyond 40mm^2, only minor reductions in the range of measurement error occur.

3.1.2 Number of Sections

Published studies of microdamage in cancellous bone use 2 to 6 sections per specimen (Table 1), and there is no clear consensus on the number of sections to analyze. The results of the simulations show that increasing the number of sections reduces the measurement error for both crack density and diffuse damage (Figure 5). However, analyzing more than three sections per specimen provides relatively little improvement on measurement error (insets of Figure 5). Although there is limited improved, the measurement error can be reduced to approximately 10% when analyzing over 200 sections of the specimen (as seen in three-dimensional imaging).

Thus, unless the measurement method is extremely inexpensive in terms of labor (i.e. three-dimensional automated approaches) analyzing more than three sections provides little benefit in terms of reducing the error.

The Effect of the Number of Sections and Bone Volume Fraction on Measurement Error for Crack Density and Diffuse Damage Volume

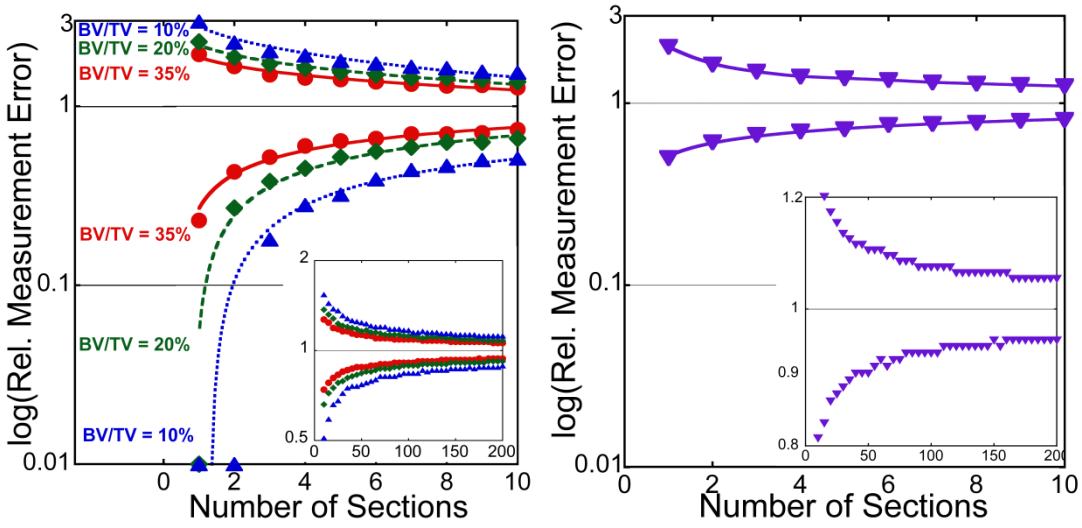

Figure 5. Increasing the number of sections reduces measurement error of crack density (left) and damage volume (right), as indicated by a thinning of the intra - 95% range. Increasing the number of sections beyond 3, however, will have little improvement on measurement error (see inset). Measurement error for diffuse damage is the same for every bone volume fraction.

3.1.3 Bone Volume Fraction

Because microdamage may only be present in bone tissue, specimens with greater bone volume fraction should have a greater amount of bone and microdamage to sample, potentially reducing measurement error. With regard to crack density, our simulations suggest that specimens with less bone volume fraction are expected to have increased measurement error as indicated by an increase in the intra-95% range of measurement error (Figures 4 and 5). However, even in more porous specimens (BV/TV = 10%) increasing the total cross-sectional area per specimen examined

beyond 120 mm^2 (3 sections of 40 mm^2 analyzed) has little effect on reducing measurement error (Figure 4). Hence, there is little advantage to examining more than 120 mm^2 of each specimen in either low-density or high-density cancellous bone. There was no effect of bone volume fraction on the measurement error of diffuse damage in our simulations.

3.1.4 Three Dimensional Measurement of Microdamage

Three-dimensional measures of microscopic tissue damage in cancellous bone have recently been demonstrated. These include the use of a radio-opaque stain such as lead-uranyl acetate [3, 10, 11] or barium sulfate [4] with micro-computed tomography and the use of standard fluorochrome stains and three-dimensional fluorescence imaging using serial milling [5, 6]. As these methods involve sampling the entire specimen, they can remove most of the variance associated with measurement technique (some measurement variance associated with image noise and thresholding will remain and is not addressed here). Increases in the number of sections to 1000 and cross-sectional area analyzed to 50.27mm^2 result in a measurement error of less than 10% (insets of Figure 5).

3.1.5 Amount of Microdamage

Martin and colleagues implied that the proportion of specimens without microdamage (‘crack-less’ specimens) in a study causes increases in measurement variance [8]. However, Martin and colleagues concentrated on naturally occurring microdamage, which is present in relatively small amounts, and did not consider

microdamage generated by controlled loading in vitro, which can be much greater in magnitude. In a Poisson distribution, the variance is equal to the mean. This implies that any increase in the mean amount of microdamage leads to a proportional increase in variance (in agreement with Martin and colleagues). However, the relative size between the mean and variance remains the same. Since measurement error in this type of study is the ratio of the estimated value over the true value, measurement error is not influenced by changes in mean microdamage. Hence, measurement error is not affected by changes to the amount of microdamage in a specimen if the same method of specimen analysis (number of sections, cross-sectional area per section, etc) remains the same. Therefore, the same sampling techniques can be used for experimental groups with low- and high- amounts of microdamage.

3.2 Model 2 – Number and Distribution of Experimental Groups

One strategy for understanding the relationship between the amount of microdamage and a predictor (such as magnitude of applied mechanical strain) is to use multiple experimental groups distributed across a range of predictor values [3, 11, 12]. A regression model is used to determine the relationship between the predictor and the amount of microdamage. Such a study design requires a selection of the number of experimental groups (where each group experiences a different amount of applied mechanical strain) and number of specimens per group. Simulations of different study designs (number of donors, number of specimens per donor, and distribution of specimens across a range of applied mechanical loading) are performed using the population distributions found in previous studies (Table 1, Table 2) and the

results from the Model 1 experiments. Two types of experiments were simulated: The first set of simulations uses multiple specimens from each donor to take advantage of repeated measures of each donor. In this first simulation model a study design with 5 experimental groups and 3 specimens per group requires only 3 donors (5 specimens from each donor for a total of 15 specimens). This first type of simulation is referred to as a "repeated measures" simulation. The second set of simulations represents an experiment where it is not possible to achieve multiple specimens from each donor and only one specimen per donor is included in the study. In this second set of simulations, a design with 5 experimental groups and 3 specimens in each group requires 15 donors (1 specimen from each donor for a total of 15 specimens). This second type of simulation is referred to as a "general" simulation.

Traditionally, statistical power is reported and used to determine the reliability of experimental results; however, more recent statistics literature has highlighted the limitations of expressing statistical power as a single number and recommends reporting a confidence interval along with the traditional statistical power [13, 14]. For this reason, both statistical power (at $\alpha = 0.05$) and the 95% confidence interval in the slope (B1) are reported for the simulations. Consistent with recent recommendations in the statistics community, a statistical power of 0.90 is considered to be acceptable for a study design [15].

3.2.1 Regression Fit Model

Two sets of simulations were performed, one where 5 experimental groups and another where 9 experimental groups were used (Table 3). Simulations for both two

and three-dimensional measures of microdamage were performed for each study design. In simulations where there are an equal number of specimens per donor as experimental groups (the ‘repeated measures’ approach), the 95% confidence interval of B_1 narrows and essential statistical power increases with an increase in the number of experimental groups (Figure 6). For example, when using two-dimensional measurement techniques with 9 donors, essential statistical power for an experiment with 5 groups (45 total specimens) is 0.8600 whereas for an experiment with 9 groups (81 total specimens) is 0.9558. When using three-dimensional measures, there is a reduction in the number of donors necessary to acquire sufficient convergence of the 95% confidence interval and essential statistical power. For an experiment with 9 groups, only 6 donors (54 total specimens, Table 3) are needed whereas for an experiment with 5 groups, only 7 donors (35 total specimens, Table 3) are needed.

Table of Experiment Designs Simulated with Results Determined by the Statistical Code

Table 2: Each simulation was performed using two and three dimensional data. If the simulation was deemed 'paired', a specimen from each donor was placed in each group (i.e. number of groups = 4, 4 specimens per donor, one specimen per group).

Experiment Type	Simulation Type	Number of Groups	Mechanical Strain Applied to Each Group	Measure	Donors per Experiment	Total Number of Specimens	Specimens per Donor
Regression Model	Repeated Measures	9	0%, 0.25% 0.5%, 0.75%, 1.0%, 1.25% 1.5%, 1.75%, 2.0%, 2.5%	2D	8	72	9
				3D	6	54	9
Regression Model	Repeated Measures	5	0%, 1.0%, 1.5%, 2.0%, 2.5%	2D	11	55	5
				3D	7	35	5
Regression Model	General	9	0%, 0.25% 0.5%, 0.75%, 1.0%, 1.25% 1.5%, 1.75%, 2.0%, 2.5%	2D	54	54	1
				3D	36	36	1
Regression Model	General	5	0%, 1.0%, 1.5%, 2.0%, 2.5%	2D	45	45	1
				3D	30	30	1
T-test	Repeated Measures	2	0%, 1%	2D	51	102	2
				3D	31	62	2
T-test	Repeated Measures	2	0%, 2.5%	2D	12	24	2
				3D	8	16	2
T-test	General	2	0%, 1%	2D	98	98	1
				3D	56	56	1
T-test	General	2	0%, 2.5%	2D	22	22	1
				3D	14	14	1

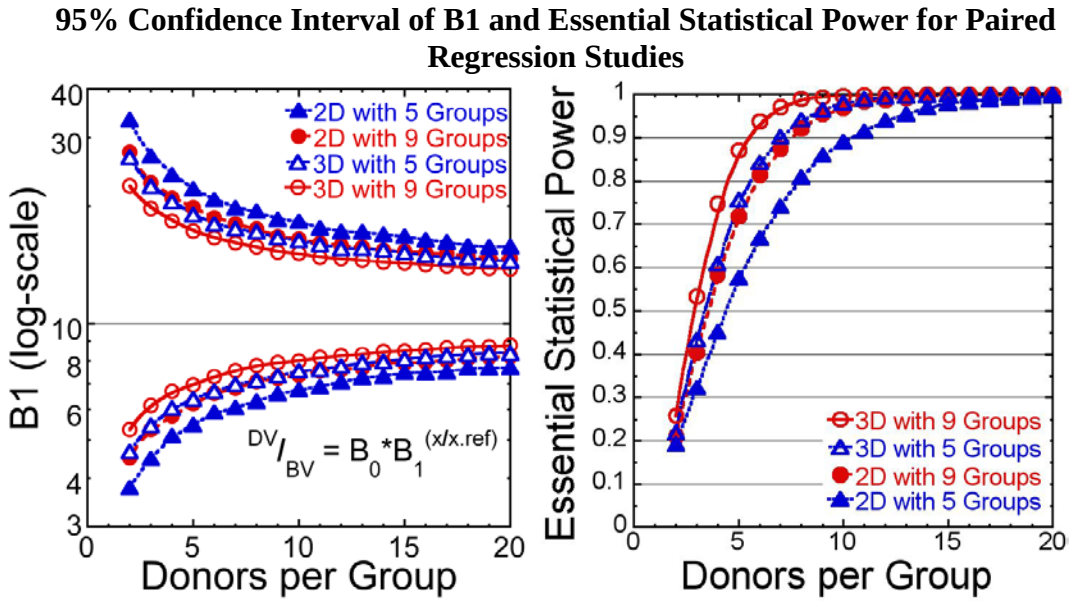

Figure 6: Simulation results of the repeated measures regression model are shown. The 95% confidence interval of B_1 (left) and essential statistical power (right) are gradually affected by the addition of donors. For example, an experiment that utilizes two-dimensional measurement techniques with 5 experimental groups would require at least 11 donors with 5 specimens per donor thus 55 total specimens.

In simulations where only one specimen per donor is used and specimens are randomly assigned experimental groups (the 'general' approach), increasing the number of experimental groups reduces the number of specimens per group required to achieve essential statistical power (Figure 7). As an example, when using two-dimensional measures of microdamage, essential statistical power of 0.90 for a study with 5 experimental groups can be achieved using 9 specimens per group (54 total specimens, Table 3). When using three-dimensional measures of microdamage even fewer specimens are required. When three-dimensional measures of microdamage are used, essential statistical power of 0.90 can be achieved when using 6 specimens per group (24 total specimens, Table 3).

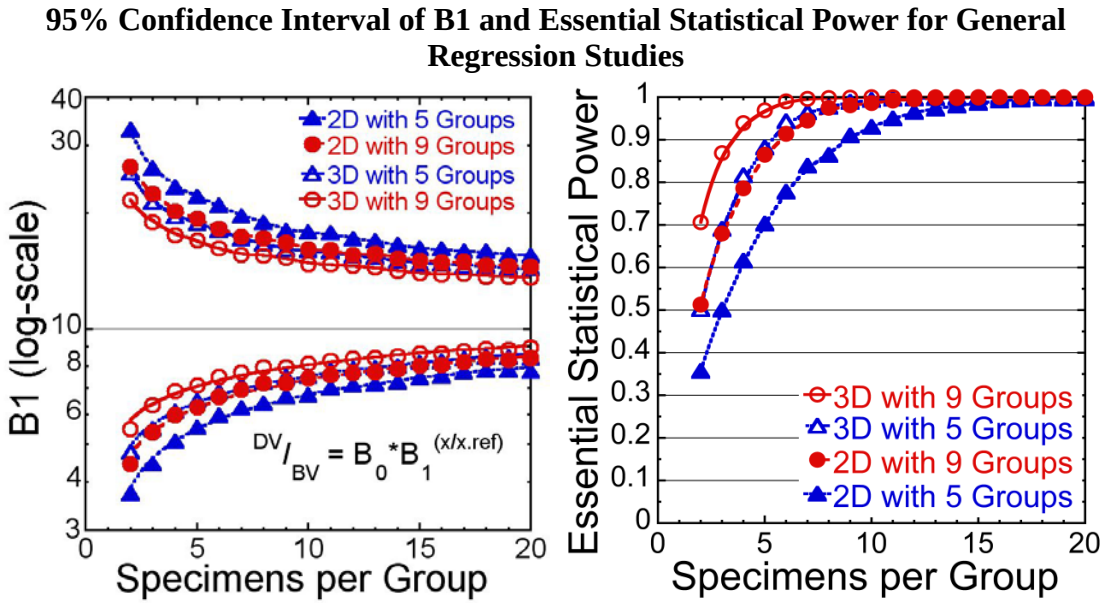

Figure 7: Simulation results using the general regression model are shown. For a general regression only one specimen per donor is used and randomly assigned an experimental group. The 95% confidence interval of B_1 (left) narrows and essential statistical power (right) increases as more specimens are added to each group. When there are a large number of experimental groups fewer specimens per group are necessary for the 95% confidence interval of B_1 and statistical power to converge than when there are fewer experimental groups. For example, an experiment that utilizes two-dimensional measurement techniques with 5 experimental groups requires at least 9 specimens per group or 45 donors, one specimen per donor.

3.2.2 Analysis of T-Tests

When only two experimental groups are included, the model is identical to a t-test (Figure 2) [9]. The most common implementation of such a design in bone biomechanics is an experiment that involves a non-loaded control group and a loaded experiment group. In the regression model the intercept of the line is the mean of the control group and the slope of the line is the size of the effect of treatment (in the current example, applied mechanical strain).

For the repeated measures approach (there are two specimens per donor, one specimen in the control group and one in the experimental group), increases in effect

size resulted in improved statistical power. If specimens are loaded to an apparent strain value just beyond yield (1.0% for human cancellous bone) and two-dimensional measures of microdamage are used, a large number of donors are needed to acquire a statistical power of 0.90 (Figure 8, Table 3). However, if the load magnitude is increased well beyond yield (apparent strain = 2.25%), the number donors of donors necessary to achieve essential statistical power using two-dimensional measurement techniques is reduced by 75% (Figure 8, Table 3). Three-dimensional measurement techniques can be used to reduce the number of donors required for essential statistical power and 95% confidence intervals of B_1 to a more manageable level (Figure 8, Table 3).

95% Confidence Interval of B1 and Essential Statistical Power for Paired T-Test Studies

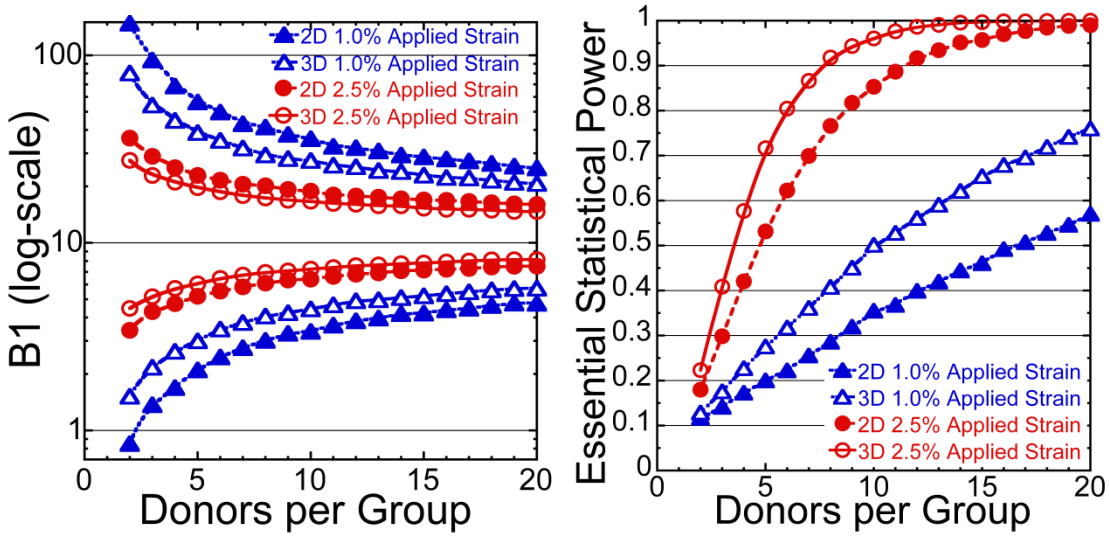

Figure 8. The 95% confidence interval of B1 (left) and essential statistical power (right) are far less sensitive to the addition of more donors than the regression models. To detect small differences in microdamage (low applied strains), a large number of donors are necessary. For example, an experiment that utilizes two-dimensional measurement techniques with an applied strain of 1.0% (a typical biomechanics study) 40 donors (80 total specimens) are required for convergence of the 95% confidence interval of B1.

For a general t-test (only one specimen per donor randomly assigned to the control or experimental group), increases in effect size resulted in improved statistical power. Changing the measurement technique from two-dimensional to three-dimensional drastically increased power and reduced the 95% confidence interval (Figure 9, Table 3).

95% Confidence Interval of B1 and Essential Statistical Power for Paired T-Test Studies

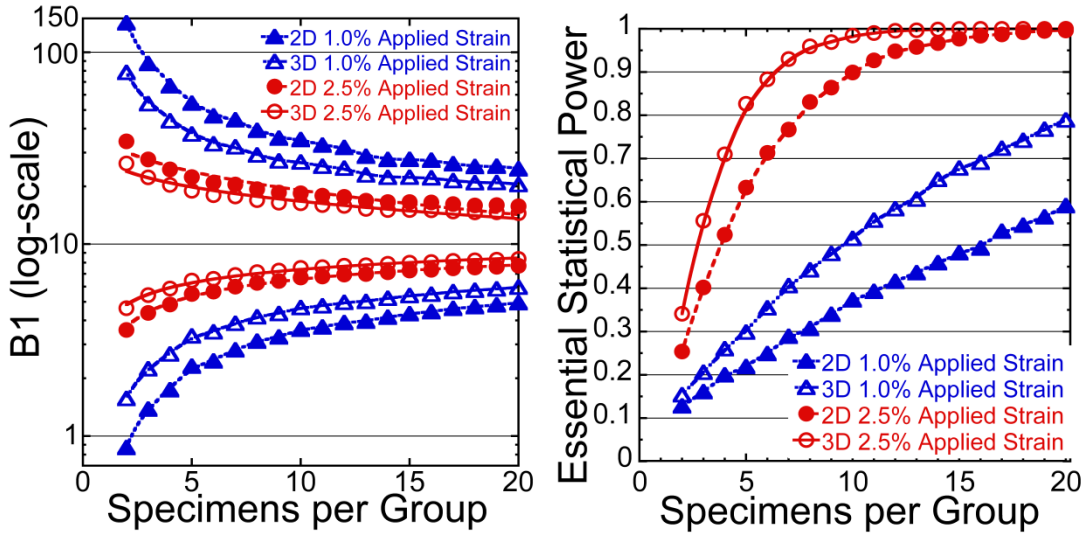

Figure 9. Experimental outcomes for a general t-test are shown. The general t-test shows very low statistical power and should only be used if large effect sizes are being detected or a large number of donors with extremely accurate and precise measurements are used. For small differences between groups (low applied strains ~1.0%), a large number of specimens per group is required. For an experiment using two-dimensional measurement techniques with an applied strain of 1.0%, 40 specimens per group (80 donors) are required for convergence of the 95% confidence interval of B_1.

4.0 DISCUSSION

The analysis suggests that when using two-dimensional stereology techniques very little benefit in measurement error is observed by examining more than 3 sections of 40 mm^2 each (Figures 4 and 5). Three-dimensional microdamage measures can further reduce measurement error and is preferred when the technique is available. Increasing the number of experimental groups will reduce the number of specimens per group needed to detect meaningful trends between applied mechanical loading and amounts of microdamage (Figures 4 and 5).

While each experiment is limited by the number of donors and complexity of specimen preparation, some ideal study designs based on this investigation are proposed. The results from the analysis suggests that when measuring microdamage using two-dimensional sections based on the sampling guidelines reported near the beginning of the paper and the number of specimens per donor is not severely limited (i.e. beams of cortical bone from the femoral diaphysis, vertebral cancellous bone, etc.) an attractive repeated measures study design uses 11 donors and 5 experimental groups where each group has one specimen from each donor (i.e. 5 specimens from each donor). In cases where the number of specimens per donor is limited (i.e. cancellous bone cores from the proximal femur, biopsy specimens, etc.) and the number of donors is not as limited a study design using 45 donors separated into 5 experimental groups using a general regression test (one specimen per donor randomly assigned to an experimental group) is recommended. Two experimental groups (t-test) are recommended only when large differences in microdamage are expected (for example when the applied strain approaches ultimate strain).

A major strength of the current study is that it generalizes previous models of sampling microdamage. The current model can be used for either cortical or cancellous bone, includes both crack density and diffuse damage measures, and incorporates both two-dimensional and three-dimensional approaches to examining microdamage. Although the implementation of this model in this thesis examines the relationship between applied loading (applied mechanical strain) and microdamage, the model design is general and can be applied to a study with any continuous and independent variable assigned as the "stimulus".

There are limitations to the analysis, however. First, the models assume that each microdamage event is independent and that each event is equally likely to occur throughout the region observed. Microdamage events may not be independent of each other, particularly when three-dimensional imaging of microdamage is used; however, the spatial relationships between sites of microscopic tissue damage are not well understood and would represent an additional unknown parameter in the statistical models.

The most efficient study design is dependent on the goal of the study. If the goal of the study is to determine the effect of a stimulus, a t-test would be sufficient; however, either the difference between the control and experimental group or the number of specimens must be large. Instead of determining if there is an effect of treatment; one may instead determine the relationship between the stimulus and response by increasing the number of experimental groups. A reasonable number of specimens, if allocated appropriately, can accurately predict the relationship between the applied stimulus and resulting amounts of microdamage. The measurement of

microdamage in bone is challenging, both from the subjective nature of identification of microdamage as well as from a sampling standpoint. Appropriate study design, however, can reduce the variance in a study and provide better insights into the formation of microdamage in bone tissue.

APPENDIX 1

Four software packages were created and are included in Appendix 2. Each file may be copied into R (available for free on most operating systems at http://www.r-project.org/). R or R Studio can be used to run any of the software packages included. To run the software, highlight the entire function within the file and 'evaluate the selection'. At the bottom of the function is an example case which then can be evaluated using the same process. Each parameter in the function can be altered due to specific experimental conditions.

The following pages include definitions of each parameter used in each of the statistical modeling software programs. A table of values used for each simulation in the manuscript is included as well.

Measurement Error Analysis

Filename: MeasurementErrorCracks.r (To be used for analysis with crack density or other parameters that are measured as a ratio of a discrete variable over a continuous variable.)

Title: Title of the simulated experiment.

S.n: Number of specimens within the simulated experiment.

M.n: Number of sections analyzed per specimen.

FieldSize: Cross-sectional area analyzed per section. The value 1 was equivalent to a cross-sectional area of 42 mm^2. Field Size is scaled based on the ratio of the desired cross-sectional area over 42.

T.median: Population median of true crack density. This parameter is utilized when the behavior of the distribution is skewed.

T.gmean: Population geometric mean of true crack density. This parameter is utilized when the measure is normally distributed.

T.RS95: The 95% relative spread of the crack density. It is the ratio of 97.5% and 2.5% quantiles. The combination of the median (T.median) and the RS_{95} (T.RS95) parameters describes the distribution of a parameter.

T.CV: The coefficient of variance defined as the Standard Deviation/Mean of crack density. The combination of the mean (T.gmean) and coefficient of variance (T.CV) completely describe a normal distribution.

BV.median: Median number of bone points counted using stereology techniques.

BV.RS95: The 95% relative spread of bone points counted using stereology techniques.

EffectSize: The effect size calculated as the ratio between the experimental and control groups for a t-test: median1/median2 or mean1/mean2. For example, if the effect size was 1.3, this would be equivalent to detecting a difference of 30% between groups.

Null: Null value of the effect size. A value of 1.0 suggests there is no difference between groups.

Alpha: Probability of a Type I error.

S.cost: The relative cost (dollars, time, etc.) of adding one more specimen to the analysis. Although not studied in the primary text, S.cost may be used to optimize effort by the observer (efficiency in measurement is an important factor of stereology).

M.cost: The relative cost (dollars, time, etc.) of adding one more section per specimen to the analysis (Not used in the study in the primary text).

PRINT: Check to either allow or suppress the printing of a summary of parameter inputs. TRUE allowed for inputs to be printed in the output window, FALSE did not allow for inputs to be printed in the output window.

Seed: A random number chosen by the user to begin the simulation. The same results to a simulation would be reported every time if the same inputs (including the seed number) were submitted.

Ntrials: Number of simulated experiments to perform.

Table of Simulation Parameters for the Microcrack Simulation Code

Table 4: Each simulation was performed using the parameters described below. Relevant figures associated with the parameters are also specified.

Parameter Name	Parameter Value	Figures Associated with Parameter
S.n	10	
M.n	1 - 200	Figure 5
FieldSize	0.714 - 1.857	Figure 4
T.median	0.003	
T.RS95	6.5	
BV.median	2000 (BV/TV = 35%) 1143 (BV/TV = 20%) 571 (BV/TV = 10%)	Figure 4 and Figure 5 (left)
BV.RS95	1.9	
EffectSize	1.3	
Null	1	
Alpha	0.05	
TestType	t-test	
S.cost	100	
M.cost	5	
PRINT	TRUE	
Seed	1234955	
Ntrials	5000	

The MeasurementErrorDiffuse.r file has a number of the same variables as the MeasurementErrorCracks.r file. Below are only the definitions of the parameters that are unique to the MeasurementErrorDiffuse.r file. A table of all the parameters used in the MeasurementErrorDiffuse.r file is also below.

Filename: MeasurementErrorDiffuse.r (To be used for analysis with diffuse damage or other parameters that are measured as a ratio of two continuous variables.)

M.RS95: The 95% relative spread of the number of damage points counted in a given section using stereology techniques. Because diffuse damage is a ratio of two continuous variables, there is a spread value that is necessary to properly define a distribution of the measurement.

M.CV: Coefficient of variation of the number of damage points counted using stereology techniques in a given section.

Table of Simulation Parameters for the Diffuse Damage Simulation Code

Table 5: Each simulation was performed using the parameters described below. Relevant figures associated with the parameters are also specified.

Parameter Name	Parameter Value	Figures Associated with Parameter
S.n	10	
M.n	1 - 200	Figure 5 (right)
FieldSize	0.714 - 1.857	Not Shown
T.median	0.003	
T.RS95	6.5	
M.RS95	4	
EffectSize	1.3	
Null	1	
Alpha	0.05	
TestType	t-test	
S.cost	100	
M.cost	5	
PRINT	TRUE	
Seed	1234955	
Ntrials	5000	

Number and Distribution of Study Groups Analysis

Filename: StudyDesignMultipleComparisons.r (To be used when there are an equal number of specimens as experimental groups per donor)

Title: Title of the simulated experiment.

AS.pts: Applied strain values for the experimental groups in the study. For example, the value 'c(0, 0.5, 1.0)' includes three experimental groups: One control with 0.0% applied compressive strain, one with 0.5% applied compressive strain, and one with 1.0% applied compressive strain. For this software, four sets of AS.pts values were used to represent four different experiment designs. Two designs represent a regression analysis with multiple applied strain values (9 experimental groups or 5 experimental groups). The other two designs represent a t-test with either a small or large difference between applied strain values.

AS.ref: The arbitrary reference strain value to properly calculate B_0 and B_1.

S.n: Number of donors in each simulated study.

M.n: Number of times each specimen is analyzed.

B0.gmean: The geometric mean value (which in this case is the same as median) of B_0 over all donors. The combination of B0.gmean and B0.RS95 completely define the distribution of B_0s for the whole population.

B0.RS95: The 95% relative spread (RS) of B_0 over all donors.

B1.gmean: The geometric mean value (which in this case is the same as median) of B_1 over all donors. The combination of B1.gmean and B1.RS95 completely define the distribution of B_1s for the whole population of interest.

B1.RS95: The 95% relative spread of B_1 over all donors.

TotalNoise.RS95: The 95% relative spread of the total randomness of each damage volume measurement. This can be determined by plotting the data, determining the regression fit, and calculating the maximum and minimum distance from the regression line.

PropMeasNoise: The proportion of the TotalNoise parameter due to measurement error. This was based off of the measurement error analysis program.

ConfLevel: The confidence level desired for the analysis.

B1.null: The null value at which we would like to compare our estimated B_1 in order to calculate statistical power. B1.null = 1 would be representative of a *typical* statistical power analysis. To be more confident in our results, we use a larger null value B1.null = 5

Print: Set to TRUE or FALSE to allow or not allow the printing of parameter data

Plot: Set to TRUE or FALSE to allow or not allow a graphical plot of simulation 1

LogScaleDV.BV: Set to TRUE to plot damage volume on a log-scale, set to FALSE to plot damage volume on a regular scale

NewWindow: Set to TRUE or FALSE to allow or not allow a new window for each plot to be generated.

Seed: A random number chosen by the user to begin the simulation. The same values would be reported by the simulation every time the same inputs (including the seed number) were submitted.

Ntrials: Number of simulated experiments to perform (5000)

Table of Simulation Parameters for the Multiple Comparisons Regression Simulation Code

Table 6: Each simulation was performed using the parameters described below. Relevant figures associated with the parameters are also specified.

Parameter Name	Parameter Value	Figures Associated with Parameter
AS.pts	c(0, 0.5, 1.0, 1.25, 1.5, 1.75, 2.0, 2.5) 9 Groups c(0,1.0, 1.5, 2.0, 2.5) 5 Groups c(0, 1.0) 2 Groups, small effect size c(0, 2.5) 2 Groups, large effect size	Figure 6 Figure 6 Figure 8 Figure 8
AS.ref	2	
S.n	2 -20	
M.n	1	
B0.gmean	0.02	
B0.RS95	3	
B1.gmean	11	
B1.RS95	2	
TotalNoise.RS95	13 (2D), 7.16 (3D)	
PropMeasNoise	0.35 (2D), 0.05 (3D)	
ConfLevel	0.95	
B1.null	5	
Print	TRUE	
Plot	TRUE	
LogScaleDV.BV	TRUE	
NewWindow	TRUE	
Seed	1234955	
Ntrials	5000	

The StudyDesignGeneral.r file has a number of the same variables as the StudyDesignMultipleComparisons.r file. Below are only the definitions of the parameters that are unique to the StudyDesignGeneral.r file. A table of all the parameters used in the StudyDesignGeneral.r file is also below.

Filename: StudyDesignGeneral.r (To be used when there only one specimen from each donor that is randomly assigned an experimental group)

n: Number of specimens in each experimental group. For example, the value 'c(2, 3, 2)' would mean that there are 2 specimens assigned to applied strain 1, 3 specimens assigned to applied strain 2, and 2 specimens assigned to applied strain 3. For all experim6ents reported here, the number of specimens in each experimental group was the same for each set of simulations (n ranges from 2 – 20) for example, n = c(18, 18, 18, 18, 18, 18, 18, 18, 18) means that there are 9 experimental groups and each group has 18 specimens in the group.

Table of Simulation Parameters for the General Regression Simulation Code

Table 7: Each simulation was performed using the parameters described below. Relevant figures associated with the parameters are also specified.

Parameter Name	Parameter Value	Figures Associated with Parameter
Title	"Description of Experiment"	
AS.pts	c(0, 0.5, 1.0, 1.25, 1.5, 1.75, 2.0, 2.5) 9 Groups c(0,1.0, 1.5, 2.0, 2.5) 5 Groups c(0, 1.0) 2 Groups, small effect size c(0, 2.5) 2 Groups, large effect size	Figure 7 Figure 7 Figure 9 Figure 9
AS.ref	2	
n	c(2, 2, 2, 2, 2, 2, 2, 2) 9 Groups c(2, 2, 2, 2, 2) 5 Groups c(2, 2) 2 Groups, small effect size c(2, 2) 2 Groups, large effect size	Figure 7 Figure 7 Figure 9 Figure 9
B0.gmean	0.02	
B0.RS95	3	
B1.gmean	11	
B1.RS95	2	
TotalNoise.RS95	13 (2D), 3.5 (3D)	
PropMeasNoise	0.35 (2D), 0.05 (3D)	
ConfLevel	0.95	
B1.null	5	
Print	TRUE	
Plot	TRUE	
LogScaleDV.BV	TRUE	
NewWindow	TRUE	
Seed	1234955	
Ntrials	5000	

APPENDIX 2

Code Written in R for Measurement Error Analysis

MeasurementErrorCracks.r

```
# BoneMeasSim.CracksPUBA
# Uses simulation to examine the statistical properties of bone measurement data.
# Ralph O'Brien, obrienralph@gmail.com
# in colloboration with Christopher Hernandez and Katherine Ehlert
# XX April 2011

# Summary.
#  S: specimen
#  S.n: number of independent specimens per group

#  M.type: type of measurement,
#         e.g. microcracks per mm^2 field, taken to stem from Poisson process.
#  M: one sample measurement such as observed number of cracks in field.
#     M varies naturally within specimen (within specimen variation)
#  M.n: number of measurements per specimen. Increasing M.n reduces measurement
#       error, shortens confidence intervals, increases power.
#  FieldSize: field size in units of measurement. Example: if unit is mm^2,
#            and field is 2mm x 2mm = 4mm^2, then FieldSize = 4.
#            Increasing FieldSize reduces measurement error, shortens confidence
#            intervals, increases power.

#  T: true value of M per unit of measurement, if the entire specimen was
#     measured perfectly. T varies from specimen to specimen (between speciment
variation)

#  Distribution of T varies over specimens depending on type of
#         measurement.
#      <> For CracksPoisson(), T varies according to a logNormal
#         distribution with paramaters
#           T.median: the median number of cracks per unit measure
#           T.RS95: the intra-95% relative spread, which is the ratio of
#                   0.975 and 0.025 quantiles. See example.
#         -or
#           T.gmean: the population geometric mean; equivalent to pop. median
#           T.CV: the coefficient of variation
#
#         Example. T.median = 0.10/mm^2 and T.RS95 = 3.0 defines 0.975 and 0.025
#                quantiles as 0.175 and 0.057.
#         Verify: round(exp(c(log(.10) - log(3)/1.96, log(.10) + log(3)/1.96)),3))
```

```
#
#         Note: Using T.median and T.RS95 is generally an easier way to characterize
#         the logNormal than is using the 'standard' parameterization based on
#         T.gmean and T.CV.


# Simulation results
# ==================

#  T.est: estimate of T for a specimen, based on all M.total measurements.
#         This is the variable used in the actual study.

#  E: error of measurement (T.est versus T) for a given specimen.
#     For example, E = T.est/T is the relative difference of T.est vs. T.
#     Such values will tend to be right skewed in distribution.
#
#  E.median: median of E over an infinite number of specimens.
#           If the measure is median unbiased, then E.median = 1.
#  E.RS95: intra-95% relative spread over infinite number of specimens, that is,
#             the ratio of the 0.975 and 0.025 quantiles.
#    Note that E.median and E.RS95 completely specify a logNormal distribution for E,
#    just as the mean and coefficient of variation do. For E.median = 1 and
#    E.RS95 = 1.2, the logNormal-based 0.025 and 0.975 quantiles for E = T.est/T
#    are 0.911 and 1.097, indicating that 95% of the estimates for T are
#    within 10% of the true values. Increasing M.total would decrease E.RS95
#    and thus move the 0.025 and 0.975 quantiles for E = T.est/T closer to
#    E.median = 1.0.
#    R code: round(exp(c(log(1) - log(1.2)/1.96, log(1) + log(1.2)/1.96)),3)

#  EffectSize: For studying a two-group comparison, this is the ratio
#             T.median[1]/T.median[2], or T.gmean[1]/T.gmean[2]
#             Example: EffectSize = 1.5 defines group 1's T.median (or T.gmean) to be
#             1.5 times that of group 2.

#  null:       Defines the test Ho: EffectSize = null
#             For this problem, the usual null is 1.0.

#  alpha:      Type I error rate. 1 - alpha is confidence level.

#  dist.T.est: distribution of estimated true values over specimens. This relates
#             to using T.est in studies. Specifically, the variability of T.est
#             stems from both within specimen and between specimen
#             variability. The former can reduced by increasing M.total. The
#             latter is set by the types of specimens being studied.
```

```
#  S.cost: # Cost per specimen, scaled to be 100.
#  M.cost: # Cost per measurement, scaled to be the % of cost of one specimen.
#          In many cases, one will use S.total specimens to estimate, say, the
#          geometric mean of T. If so, then the standard error of that estimate
#          will be reduced by increasing either M.total or S.total. If the total
#          cost is limited to C = M.cost*M.total + S.cost*S.total, what is the
#          optimal choice of M.total and S.total that results in the lowest
#          stanaard error?

CracksPUBA = function(
  title="a case with no title",
  S.n,            # number of specimens
  M.n,             # number of measurements per specimen
  FieldSize=1,     # number of units of measure per field
  T.median=NA,     # pop. median of true crack densities, T, for group 1, per unit
measure
  T.gmean=NA,      # pop. geometric mean of T, exp(mean(log(T))); equals pop.
median T
  T.RS95=NA,       # 95% relative spread of T; ratio of T(0.975) and T(0.025)
quantiles
  T.CV=NA,         # coeffcient of variation, SD/mean
  BV.median,       # median BV
  BV.RS95,         # 95% relative spread of T
  EffectSize=1.0,  # effect size, median1/median2 or gmean1/gmean2
  null=1.0,        # null value for effect size
  alpha=0.05,      # Type I error rate; 1 - alpha is confidence level
  TestType="t.test",# "t.test" = Welch t test on log(M) (var.equal = FALSE)
                # "GLM" Poisson regression, to be added
  S.cost=100,      # cost per specimen
  M.cost=NA,       # cost per measure
  PRINT=TRUE,      # set to FALSE to suppress printing
  seed=1234955,    # a seed to start simulations
  Ntrials=5000)    # number of trials in simulations
 {

 if (PRINT) {
  cat("\n\n",title,"\n\n")
  print(data.frame(S.n, M.n,FieldSize),row.names=""); cat("\n")
  print(data.frame(T.median,T.RS95,T.gmean,T.CV),row.names=""); cat("\n")
  print(data.frame(EffectSize,null,alpha,TestType,row.names="")); cat("\n")
  print(data.frame(S.cost,M.cost,row.names=""))
  cat("\nTotal Cost:", S.cost*S.n + M.cost*M.n,"\n\n")
  print(data.frame(PRINT,seed,Ntrials,row.names="")); cat("\n")
 }
```

```
set.seed(seed)
T <- rep(NA,Ntrials)
T.est <- rep(NA,Ntrials)
M <- rep(NA,S.n)
group <- c(rep(0,S.n), rep(1,S.n))
EffSize <- c(rep(1,S.n), rep(1/EffectSize,S.n))
estimate <- rep(NA,Ntrials)
CI.lo <- rep(NA,Ntrials)
CI.hi <- rep(NA,Ntrials)
BV <- rep(NA,Ntrials)

if (!is.na(T.median)) {
  if (T.median <= 0) {
    stop ("\n\nT.median must be positive.")
  }
  EffectType <- "T.median[1]:T.median[2]"
  mean.logT <- log(T.median)
} else if (!is.na(T.gmean)) {
    if (T.gmean <= 0) {
      stop ("\n\nT.gmean must be positive.")
    }
    EffectType <- "T.gmean[1]:T.gmean[2]"
    mean.logT <- log(T.gmean)
  } else {
      stop ("\n\nMust supply either T.median or T.gmean.")
  }

if (!is.na(T.RS95)) {
  if (T.RS95 <= 0) {
    stop ("\n\nT.RS95 must be positive.")
  }
  SD.logT <- log(T.RS95)/(1.96*2)
} else if (!is.na(T.CV)) {
    if (T.CV <= 0) {
      stop ("\n\nT.CV must be positive.")
    }
    SD.logT <- sqrt(log(T.CV^2 + 1))
  } else {
      stop ("\n\nMust supply either T.RS95 or T.CV.")
  }

meanLnBV <- log(BV.median)
StdevLnBV <- log(BV.RS95)/(2*1.96)
```

```
  for (itrial in 1:Ntrials)  {
    BV[itrial] <- round(exp(rnorm(1,meanLnBV,StdevLnBV)))
    T[itrial] <- exp(rnorm(1,mean.logT,SD.logT))
    T.est[itrial] <-
mean(rpois(M.n,T[itrial]*BV[itrial]*FieldSize))/(BV[itrial]*FieldSize)

    for (S in 1:(2*S.n)) {
      BV.S <- round(exp(rnorm(1,meanLnBV,StdevLnBV)))
      T.S <- exp(rnorm(1,mean.logT,SD.logT))  # T for this S
      M[S] <- mean(rpois(M.n,EffSize[S]*T.S*BV.S*FieldSize))/(BV.S*FieldSize)
      if (M[S] == 0) {
        cat("\n\nNote: A specimen was observed to have O total cracks over all
measurements",
            "\nThis was set to be 0.5 total cracks.\n")
        M[S] <- (0.50/M.n)/(BV.S*FieldSize)
      }
    }

    if (TestType == "t.test") {
      tt <- t.test(log(M)~group, conf.level=1-alpha)
      estimate[itrial] <- exp(tt$estimate[1])/exp(tt$estimate[2])
      CI.lo[itrial] <- exp(tt$conf.int[1])
      CI.hi[itrial] <- exp(tt$conf.int[2])
    }
  }
  E <- T.est/T
  E.qvals <- quantile(E,c(0.025,0.500,0.975))

  SD.logT.est <- sd(T.est)
  SE.logT.est <- SD.logT.est/sqrt(S.n)

  CLimsRatio <- CI.hi/CI.lo
  CLims.qvals <- quantile(CLimsRatio,c(0.025,0.500,0.975))
  power <- sum((CI.lo > null) + (CI.hi < null) > 0)/Ntrials

  if (PRINT) {
    cat("\n\nDistribution of E = estimate:actual")
    cat  ("\n================================")
    cat("\nMedian:", round(E.qvals[2],3))
    cat("\n0.025 & 0.975 quantiles:", round(E.qvals[1],2), round(E.qvals[3],2))
    cat("\n95% relative spread:", round(E.qvals[3]/E.qvals[1],2),"\n")

    cat("\nProperties of", round(1-alpha,3), "confidence interval for", EffectType)
```

```
   cat
("\n=============================================================
=======")
   cat("\nMedian lower and upper limits:",round(median(CI.lo),2),
round(median(CI.hi),2))
   cat("\nMedian ratio upper:lower limits:",round(median(CI.hi/CI.lo),2))
   cat("\n\nProportion not containing null =", null,":", round(power,3))
   cat ("\n(This is the statistical power of this case.)\n")
  }
# results <= list
# invisible(qvals)
}
```

MeasurementErrorDiffuse.r

```
# MeasurementErrorCracks.r
# Uses simulation to examine the statistical properties of bone measurement data.
# Ralph O'Brien, obrienralph@gmail.com
# in collaboration with Christopher Hernandez and Katherine Ehlert
# XX April 2011

# Summary.
#  S: specimen
#  S.n: number of independent specimens per group

#  M.type: type of measurement,
#         e.g. microcracks per mm^2 field, taken to stem from Poisson process.
#  M: one sample measurement such as observed number of cracks in field.
#     M varies naturally within specimen (within specimen variation)
#  M.n: number of measurements per specimen. Increasing M.n reduces measurement
#       error, shortens confidence intervals, increases power.
#  FieldSize: field size in units of measurement. Example: if unit is mm^2,
#            and field is 2mm x 2mm = 4mm^2, then FieldSize = 4.
#            Increasing FieldSize reduces measurement error, shortens confidence
#            intervals, increases power.

#  T: true value of M per unit of measurement, if the entire specimen was
#     measured perfectly. T varies from specimen to specimen (between speciment
variation)

#  Distribution of T varies over specimens depending on type of
#         measurement.
#      <> For CracksPoisson(), T varies according to a logNormal
#         distribution with paramaters
#           T.median: the median number of cracks per unit measure
#           T.RS95: the intra-95% relative spread, which is the ratio of
```

```
#                   0.975 and 0.025 quantiles. See example.
#        -or
#          T.gmean: the population geometric mean; equivalent to pop. median
#          T.CV: the coefficient of variation
#
#        Ezxample. T.median = 0.10/mm^2 and T.RS95 = 3.0 defines 0.975 and 0.025
#               quantiles as 0.175 and 0.057.
#        Verify: round(exp(c(log(.10) - log(3)/1.96, log(.10) + log(3)/1.96)),3))
#
#        Note: Using T.median and T.RS95 is generally an easier way to characterize
#        the logNormal than is using the 'standard' parameterization based on
#        T.gmean and T.CV.


# Simulation results
# ==================

#  T.est: estimate of T for a specimen, based on all M.total measurements.
#        This is the variable used in the actual study.

#  E: error of measurement (T.est versus T) for a given specimen.
#    For example, E = T.est/T is the relative difference of T.est vs. T.
#    Such values will tend to be right skewed in distribution.
#
#  E.median: median of E over an infinite number of specimens.
#          If the measure is median unbiased, then E.median = 1.
#  E.RS95: intra-95% relative spread over infinite number of specimens, that is,
#            the ratio of the 0.975 and 0.025 quantiles.
#   Note that E.median and E.RS95 completely specify a logNormal distribution for E,
#   just as the mean and coefficient of variation do. For E.median = 1 and
#   E.RS95 = 1.2, the logNormal-based 0.025 and 0.975 quantiles for E = T.est/T
#   are 0.911 and 1.097, indicating that 95% of the estimates for T are
#   within 10% of the true values. Increasing M.total would decrease E.RS95
#   and thus move the 0.025 and 0.975 quantiles for E = T.est/T closer to
#   E.median = 1.0.
#   R code: round(exp(c(log(1) - log(1.2)/1.96, log(1) + log(1.2)/1.96)),3)

#  EffectSize: For studying a two-group comparison, this is the ratio
#            T.median[1]/T.median[2], or T.gmean[1]/T.gmean[2]
#            Example: EffectSize = 1.5 defines group 1's T.median (or T.gmean) to be
#            1.5 times that of group 2.

#  null:      Defines the test Ho: EffectSize = null
#            For this problem, the usual null is 1.0.
```

```
#  alpha:      Type I error rate. 1 - alpha is confidence level.

#  S.cost: # Cost per specimen, scaled to be 100.
#  M.cost: # Cost per measurement, scaled to be the % of cost of one specimen.
#           In many cases, one will use S.total specimens to estimate, say, the
#           geometric mean of T. If so, then the standard error of that estimate
#           will be reduced by increasing either M.total or S.total. If the total
#           cost is limited to C = M.cost*M.total + S.cost*S.total, what is the
#           optimal choice of M.total and S.total that results in the lowest
#           stanaard error?

DV_BV = function(
  title="a case with no title",
  S.n,            # number of specimens
  M.n,             # number of measurements per specimen
  FieldSize=1,      # number of units of measure per field
  T.median=NA,      # pop. median of DV/BV, percent of bone with damage over
entire specimen
  T.gmean=NA,       # pop. geometric mean of T, exp(mean(log(T))); equals pop.
median T
  T.RS95=NA,        # 95% relative spread of T; ratio of T(0.975) and T(0.025)
quantiles
  T.CV=NA,          # coeffcient of between-specimen variation, SD/mean
  M.RS95=NA,        # 95% relative spread of M within specimens
                # M.RS95 is restricted to be <70.1% of T.RS95
                # Default: M.RS95 = T.RS95/2 (50%)
  M.CV=NA,          # coeffcient of within-specimen variation
                # M.CV is restricted to be <70.1% of T.CV
  EffectSize=1.0,   # effect size, median1/median2 or gmean1/gmean2
  null=1.0,         # null value for effect size
  alpha=0.05,       # Type I error rate; 1 - alpha is confidence level
  TestType="t.test",# "t.test" = Welch t test on log(M) (var.equal = FALSE)
  S.cost=100,       # cost per specimen
  M.cost=NA,        # cost per measure
  PRINT=TRUE,       # set to FALSE to suppress printing
  seed=1234955,     # a seed to start simulations
  Ntrials=5000)     # number of trials in simulations
  {

 if (PRINT) {
   cat("\n\n",title,"\n\n")
   print(data.frame(S.n, M.n,FieldSize),row.names=""); cat("\n")
   print(data.frame(T.median,T.RS95,T.gmean,T.CV),row.names=""); cat("\n")
   print(data.frame(EffectSize,null,alpha,TestType,row.names="")); cat("\n")
   print(data.frame(S.cost,M.cost,row.names=""))
```

```
  cat("\nTotal Cost:", S.cost*S.n + M.cost*M.n,"\n\n")
  print(data.frame(PRINT,seed,Ntrials,row.names="")); cat("\n")
}

set.seed(seed)
T <- rep(NA,Ntrials)
T.est <- rep(NA,Ntrials)
M <- rep(NA,S.n)
group <- c(rep(0,S.n), rep(1,S.n))
EffSize <- c(rep(1,S.n), rep(1/EffectSize,S.n))
estimate <- rep(NA,Ntrials)
CI.lo <- rep(NA,Ntrials)
CI.hi <- rep(NA,Ntrials)

if (!is.na(T.median)) {
  if (T.median <= 0) {
    stop ("\n\nT.median must be positive.")
  }
  EffectType <- "T.median[1]:T.median[2]"
  mean.logT <- log(T.median)
} else if (!is.na(T.gmean)) {
    if (T.gmean <= 0) {
      stop ("\n\nT.gmean must be positive.")
    }
    EffectType <- "T.gmean[1]:T.gmean[2]"
    mean.logT <- log(T.gmean)
  } else {
      stop ("\n\nMust supply either T.median or T.gmean.")
  }

if (!is.na(T.RS95)) {
  if (T.RS95 <= 0) {
    stop ("\n\nT.RS95 must be positive.")
  }
  SD.logT <- log(T.RS95)/(1.96*2)
} else if (!is.na(T.CV)) {
    if (T.CV <= 0) {
      stop ("\n\nT.CV must be positive.")
    }
    SD.logT <- sqrt(log(T.CV^2 + 1))
  } else {
      stop ("\n\nMust supply either T.RS95 or T.CV.")
  }

if ((is.na(M.RS95)) & (is.na(M.CV))) {
```

```
    M.RS95 -> exp(2*1.96*SD.logT)/2
    if (PRINT) {
      cat ("\n\nNote: M.RS95 has been set to 50% of S.RS95.")
    }
  }

  if (!is.na(M.RS95)) {
    if (M.RS95 <= 0) {
      stop ("\n\nM.RS95 must be positive.")
    }
    if (M.RS95 > .70*exp(2*1.96*SD.logT)) {
      stop ("\n\nM.RS95 must not exceed 70% of S.RS95.")
    }
    SD.logM <- log(M.RS95)/(1.96*2)
  } else {
      if (M.CV <= 0) {
        stop ("\n\nM.CV must be positive.")
      }
    if (M.CV > .70*sqrt(exp(SD.logT))) {
      stop ("\n\nM.CV must not exceed 70% of S.CV.")
    }
      SD.logM <- sqrt(log(M.CV^2 + 1))
    }

  for (itrial in 1:Ntrials)  {
    T[itrial] <- exp(rnorm(1,mean.logT,SD.logT))
    T.est[itrial] <- exp(mean(rnorm(M.n,log(T[itrial]),SD.logM))) # gmean of M.n
measurements

    for (S in 1:(2*S.n)) {
      T.S <- exp(rnorm(1,mean.logT,SD.logT))  # T for this S
      M[S] <- exp(mean(rnorm(M.n,log(EffSize[S]*T.S),SD.logM))) # gmean of M.n
measurements
    }

    if (TestType == "t.test") {
      tt <- t.test(log(M)~group, conf.level=1-alpha)
      estimate[itrial] <- exp(tt$estimate[1])/exp(tt$estimate[2])
      CI.lo[itrial] <- exp(tt$conf.int[1])
      CI.hi[itrial] <- exp(tt$conf.int[2])
    }
  }
  E <- T.est/T
  E.qvals <- quantile(E,c(0.025,0.500,0.975))
```

```
  SD.logT.est <- sd(T.est)
  SE.logT.est <- SD.logT.est/sqrt(S.n)

  CLimsRatio <- CI.hi/CI.lo
  CLims.qvals <- quantile(CLimsRatio,c(0.025,0.500,0.975))
  power <- sum((CI.lo > null) + (CI.hi < null) > 0)/Ntrials

  if (PRINT) {
    cat("\n\nDistribution of E = estimate:actual")
    cat  ("\n==================================")
    cat("\nMedian:", round(E.qvals[2],3))
    cat("\n0.025 & 0.975 quantiles:", round(E.qvals[1],2), round(E.qvals[3],2))
    cat("\n95% relative spread:", round(E.qvals[3]/E.qvals[1],2),"\n")

    cat("\nProperties of", round(1-alpha,3), "confidence interval for", EffectType)
    cat
("\n============================================================
=======")
    cat("\nMedian lower and upper limits:",round(median(CI.lo),2),
round(median(CI.hi),2))
    cat("\nMedian ratio upper:lower limits:",round(median(CI.hi/CI.lo),2))
    cat("\n\nProportion not containing null =", null,":", round(power,3))
    cat ("\n(This is the statistical power of this case.)\n")
  }
# results <= list
# invisible(qvals)
}
```

Code Written in R for Number and Distribution of Study Groups Analysis

StudyDesignPaired.r

```
# BoneMeasSimDV_BVcurve_____.r
# Uses simulation to examine the statistical properties of bone measurement data.
# Ralph O'Brien, obrienralph@gmail.com
# in collaboration with Christopher Hernandez and Katherine Ehlert
# 2 May 2011

# Background.
# Figure 4 in Tang and Vashishth (2007) plots results for their study of a single
specimen,
# (S.n = 1), which measured and fit Damaged Bone Fraction (DV/BV, here DV.BV) to
# % Apparent Strain (AS) at 9 design points:
#     AS.pts <- c(0, 0.5, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00, 2.25)
# It appears that only one measurement (M.n = 1) was taken at each design point.
# The multiplicative exponential model fit in that study was
#    [M1]  DV.BV ~ 0.02 * exp(1.2*AS)
# or, equivalently, the log-linear model
#    [L1]  log(DV.BV) ~ log(0.02) + 1.2*AS

# Aim.
# This program performs simualations that generalize this type of study by generating
# and analyzing data that conform to the model
#    [M2]  DV.BV ~ B0 * B1^(AS/AS.ref)
# where:
#       B0 = gmean(AS=0), the geometric mean (here, same as median) of DV/BV
when
#            % Apparent Strain = 0;
#       B1 = gmean(AS.ref)/gmean(AS=0), the fold change in DV/BV when
#            going from AS = 0 to some reference point for AS, AS.ref > 0.
#
# The log-linear model is
#     [L2]  log(DV.BV) ~ log(B0) + (AS/AS.ref) * log(B1)

# Note. It is straightforward to show that M2 amd M1 are equivalent models, just
expressed
# in a different ways:
#     [M1] DV.BV ~ B0 * exp(B1'*AS) = B0 * exp((ref.AS*B1')*(AS/AS.ref)) =
#                B0 + B1^(AS/AS.ref),
# where B1 = exp(ref.AS*B1'). The parameters in model [M2] are simpler to use.

# For example, take the Tang and Vashishth (2007) plot. At AS = 0 and ref.AS = 2,
# DV/BV values are about 0.02 and 0.22, so the model is roughly
```

```
#           DV.BV ~ 0.02 * (0.22/0.02)*(AS/2) = 0.02 * 11^(AS/2),
# saying that DV/BV is 0.02 at 0% Apparent Strain, 0.02*11^(0.50) = 0.02*sqrt(11) =
# 0.066 at AS = 1%, and 0.02*11 = 0.22 at AS = 2%.
#
# Note. This is nearly identical to the fit shown in T&V's Figure 4:
#      DV.BV = 0.02 * exp(1.2*AS) = 0.02 * exp(2.4)^(AS/2) = 0.02 * 11.02*(AS/2).
#
# Specifically, this function allows for varying the
#   (1) number and position of the AS design points. Note. Using two design points,
AS=0 vs AS=ref.AS,
# sets up a two-group problem, which is handled here by a simple t-test whether the
geometric mean of
# [DV.BV(AS=ref.AS)]/[DV.BV(AS=0)] = 1.0. See example.
#   (2) values for gmeans at each design point.
#   (3) relative variation of "true" values for B0 and B1 from specimen to specimen.
#   (4) relative variation of observed values of DV.BV from measurement to
measurement within
#       each subject.
#   (5) number of specimens; each specimen measured M.n times at each design point.
#   (6) number of measurements per specimen per design point.
#
# All relevant parameter values are modifiable in the function call; see examples.


# Summary.
# ========
#     AS.pts: Apparent Strength design points, such as
#           AS.pts = c(0, 0.5, 0.75, 1.00, 1.25 1.50, 1.75, 2.00, 2.25)

#     ref.AS: reference design point, single value, such as ref.AS = 2.

#       S.n: number of independent specimens. Each specimen is measured at each
#            design point.

#       M.n: number of measurements at each design point.

#     B0.gmean: geometric mean value (same as median) of B0 over specimens. As
above,
#             B0 is the true value of DV/BV when % Apparent Strain = 0. B0 varies
#             from specimen to specimen. Assumed to be logNormal in shape.

#      B0.RS95: the 95% relative spread of B0 over specimens. For example,
#             B0.RS95 = 3 indicates that the top versus the bottom of the middle 95%
#             of the B0 logNormal distribution has a ratio of 3.0.
#             Note: B0 is not varying within specimen; such randomness has been
```

```
#             dumped into the noise component.
#
#    B1.gmean: geometric mean value (same as median) of B1 over specimens. As
above,
#             B1 is the median(ref.AS)/median(AS=0), the fold change in DV/BV when
#             going from AS = 0 to some reference point for AS, ref.AS > 0. B1 varies
#             from specimen to specimen. Assumed to be logNormal in shape.

#     B1.RS95: the 95% relative spread of B1 over specimens. For example,
#             B0.RS95 = 2 indicates that the top versus the bottom of the middle 95%
#             of the B1 logNormal distribution has a ratio of 2.0.
#             Note: B1 is not varying within specimen; such randomness has been
#             dumped into the noise component.

# TotalNoise.RS95:
#             the 95% relative spread of the total randomness of each DV/BV
#             measurement. This is due to variation within the specimen
#             and measurement error.
#
# PropMeasNoise:
#            proportion of TotalNoise due to measurement error
#
#    ConfLevel: Sets level of confidence intervals (e.g. 0.95).
#
#     B1.null: sets null for testing B1. The usual value is 1.0. "Rejecting"
#             this says there is some relationship between AS and DV/BV. So what?


# Simulation results, per trial
# ============================

#   B1.est: geometric mean of B1 estimated over the specimens.

#    B1.lo: lower 95% confidence limit for B1.
#    B1.hi: upper 95% confidence limit for B1.
#
#  p.value: one-sided p.value for testing
#            Ho: B1.gmean = B1.null
#            Ha: B1.gmean > B1.null


DV.BVcurve = function(
  title="a case with no title",
  AS.pts,          # design points for Apparent Strain,
  AS.ref,          # reference design point for Apparent Strain,
```

```
  S.n,              # number of specimens
  M.n,              # number of measurements per specimen at each design point
  B0.gmean,         # geometric mean value (same as median) of B0 over specimens.
  B0.RS95,          # the 95% relative spread (RS) of B0 over specimens.
  B1.gmean,         # geometric mean value (same as median) of B1 over specimens.
  B1.RS95,          # the 95% RS of B1 over specimens.
  TotalNoise.RS95,   # the 95% RS of the total randomness of each DV/BV
measurement.
  PropMeasNoise=0.3, # proportion of TotalNoise due to measurement error
  ConfLevel=0.95,    # level of confidence interval
  B1.null,          # sets null for testing B1
  Print=TRUE,        # set to FALSE to suppress printing
  Plot=TRUE,         # set to FALSE to suppress plotting of Trial 1 data
  LogScaleDV.BV=F,   # set to TRUE to log scale DV/BV axis
  NewWindow=TRUE,    # set to FALSE to re-use the same plot window
  seed=1234955,      # a seed to start simulations
  Ntrials=1000)      # number of trials in simulations
 {

 if (Print) {
  cat("\n\n",title,"\n")
  cat("\nDesign Points:", AS.pts)
  cat("\nRefefence Point:", AS.ref, "\n")
  print(data.frame(S.n, M.n),row.names=""); cat("\n")
  print(data.frame(B0.gmean,B0.RS95,B1.gmean,B1.RS95),row.names="")
  cat("\n")
  print(data.frame(TotalNoise.RS95, PropMeasNoise), row.names="")
  cat("\n")
  print(data.frame(ConfLevel,B1.null,row.names="")); cat("\n")
  print(data.frame(Print,seed,Ntrials,row.names="")); cat("\n")
  print(data.frame(Plot,LogScaleDV.BV,NewWindow,row.names="")); cat("\n")

 }

 set.seed(seed)
 B1.est <- B1.lo <- B1.hi <- p.value <- CIspan<- rep(NA,Ntrials)
 K <- length(AS.pts)

 if (min(c(B0.gmean, B1.gmean)) <= 0) {
  stop ("B0.gmean and B1.gmean must be positive.")
  }

 if (min(c(B0.RS95, B1.RS95)) <= 1) {
  stop ("B0.RS95 and B1.RS95 must exceed 1.0.")
 }
```

```
# Note: model assumes independence between B0 and B1. This can be changed.
  mu.logB0 <- log(B0.gmean)
  mu.logB1 <- log(B1.gmean)
  SD.logB0 <- log(B0.RS95)/(1.96*2)
  SD.logB1 <- log(B1.RS95)/(1.96*2)
  SD.logTNoise <- log(TotalNoise.RS95)/(1.96*2)

# Note: model assumes independence among noise values for all measures between
# design pts. This can be changed. But assumes that measurements within design
points are
# correlated such that PropMeasNoise of the total noise variance (not SD) is
measurement
# error within design points. Probably impossible to know in any planning process.
  SDnoise.meas <- sqrt((SD.logTNoise^2)*PropMeasNoise)
  SDnoise.other <- sqrt((SD.logTNoise^2)*(1-PropMeasNoise))

  DV.BV <- matrix(rep(NA, S.n*M.n*K), nrow=S.n*M.n)
  colnames(DV.BV) <- paste("DV.BV", 1:K, sep="")
  donor=rep(1:S.n, each=M.n)
  m=rep(1:M.n, times=S.n)

  for (iTrial in 1:Ntrials)  {
    logB1.s <- rep(NA, S.n)

    ii <- 0
    for (s in 1:S.n) {
      logDV.BV <- X.AS <- rep(NA, K*M.n)

      logB0 <- rnorm(1, mu.logB0, SD.logB0)
      logB1 <- rnorm(1, mu.logB1, SD.logB1)

      i <- 0
      for (k in 1:K) {
        AS <- AS.pts[k]
        tempDV.BV <- logB0 + (AS/AS.ref)*logB1 + rnorm(1,0,SDnoise.other)
        for (m in 1:M.n) {
          i <- i + 1
          X.AS[i] <- AS
          logDV.BV[i] <- tempDV.BV + rnorm(1,0,SDnoise.meas)
          DV.BV[ii+m,k] <- exp(logDV.BV[i])
        }
      }
      ii <- ii + M.n
```

```
    if (Print & (iTrial==1) & (s==1)) {
      cat("\n\nFirst specimen in first trial\n")
      print (data.frame(AS=X.AS, DV.BV=round(exp(logDV.BV),3)))
    }

    X.ASvRef <- X.AS/AS.ref
    logB1.s[s] <- lm(logDV.BV ~ X.ASvRef)$coefficients[2]
  }

  if (Plot & (iTrial==1)) {  # make spaghetti plot of Trial 1 data
    plottitle <- paste("Data for Trial #1\nDonors:", S.n, "    Measurements per Donor:",
M.n)
                                        if (NewWindow) {
      try(windows(width=6, height=5), silent=T)
      try(quartz(width=6, height=5), silent=T)
    }

Ymin <- 0; LogScaling = ""; Ylabel <- "Damaged Volume Fraction (DV/BV)"
if (LogScaleDV.BV) {
LogScaling = "y"
DV.BV[DV.BV[,] < 0.001] = 0.001
Ymin <- min(0.001, min(DV.BV))
Ylabel <- "Damaged Volume Fraction (DV/BV; log-scaled)"
}
Ymax <- max(0.8, max(DV.BV))
plot(DV.BV[1,] ~ AS.pts,
ylim= c(Ymin, Ymax),
xlim= c(0, 2.5),
ylab=Ylabel,
xlab = "Apparent Strain (%)",
type="l", lwd=0.8, col=1,
xaxt="n",
log=LogScaling,
las=1,
cex.axis= 0.75, cex.lab=0.85,
main=plottitle)
axis(1, at=c(0, 0.5, 1, 1.5, 2, 2.5), cex.axis=0.75)
       for (i in 2:length(DV.BV[,1])) {
                        lines(DV.BV[i,] ~ AS.pts, col=donor[i], lwd=0.8)
       }
  }

  null <- log(B1.null)
  t <- t.test(logB1.s, mu=null, conf.level=ConfLevel)
```

```
    names(t)
    B1.est[iTrial] <- exp(t$estimate)
    B1.lo[iTrial] <- exp((t$conf.int[1]))
    B1.hi[iTrial] <- exp((t$conf.int[2]))
    CIspan[iTrial] <- B1.hi[iTrial]/B1.lo[iTrial]
    if (B1.est[iTrial] > 1) {
      p.value[iTrial] <- t$p.value/2
    }
      else {
       p.value[iTrial] <- 1 - t$p.value/2
    }

  }

  power <- sum((p.value < 0.05) > 0)/Ntrials
  p.qvals <- quantile(p.value,c(0.25,0.500,0.75,0.80,power))
  B1.qvals <- quantile(B1.est,c(0.025,0.975))

  if (Print) {
  cat("\n\nResults for first 10 trials\n")
  first10 <- data.frame(B1.est=round(B1.est,2),
                        B1.lo=round(B1.lo,2),
                        B1.hi=round(B1.hi,2),
                        p.value=p.value,
                        CIspan=round(CIspan,2))[1:10,]
  print(first10)
  }

# Preliminary code to analyze the simulation study
{
  cat("\n\nSummary of B1 estimates; true =", B1.gmean, "\n")
  print(round(exp(summary(log(B1.est))),2))
  cat("\n\n95% quantile range of the B1 estimates\n")
  print(round(B1.qvals,2))
  cat("\n\nSummary of B1 lower confidence limits\n")
  print(round(exp(summary(log(B1.lo))),2))
  cat("\n\nSummary of B1 upper confidence limits\n")
  print(round(exp(summary(log(B1.hi))),2))
  cat("\n\nSummary of CI span = B1.hi/B1.lo\n")
  print(round(exp(summary(log(CIspan))),2))
  cat("\n\nSummary of B1 p.values; one-sided test of B1 > ", B1.null, "\n")
  print(summary(p.value,3))
  cat("\n Quantiles\n 0.25   0.50   0.75   0.80   power\n", round(p.qvals[1],4),
round(p.qvals[2],4), round(p.qvals[3],4), round(p.qvals[4],4), round(p.qvals[5],4))
  cat("\n\n Proportion with a p.value < 0.05 :", round(power,4))
```

```
  cat("\n(This should be the statistical power of this case.)\n")
  }

} # end DV.BVcurve ()
```

StudyDesignGeneral.r

```
# OneSpecPerDonorDV_BVcurve_____.r
# Uses simulation to examine the statistical properties of bone measurement data study
# in which only one specimen is used per donor.
# Ralph O'Brien, obrienralph@gmail.com
# in collaboration with Christopher Hernandez and Katherine Ehlert
# 7 Sept 2012

# Background.
# Figure 4 in Tang and Vashishth (2007) plots results for their study of a single
specimen,
# (S.n = 1), which measured and fit Damaged Bone Fraction (DV/BV, here DV.BV) to
# % Apparent Strain (AS) at 9 design points:
#     AS.pts <- c(0, 0.5, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00, 2.25)
# It appears that only one measurement (M.n = 1) was taken at each design point.
# The multiplicative exponential model fit in that study was
#    [M1]  DV.BV ~ 0.02 * exp(1.2*AS)
# or, equivalently, the log-linear model
#    [L1]  log(DV.BV) ~ log(0.02) + 1.2*AS

# Aim.
# This program performs simulations that alter this type of study in that only one
specimen
# is used from each donor.
# We use the statistical model
#    [M2]  DV.BV ~ B0 * B1^(AS/AS.ref)
# where:
#       B0 = gmean(AS=0), the geometric mean (here, same as median) of DV/BV
when
#            % Apparent Strain = 0;
#       B1 = gmean(AS.ref)/gmean(AS=0), the fold change in DV/BV when
#            going from AS = 0 to some reference point for AS, AS.ref > 0.
#
# The log-linear model is
#     [L2]  log(DV.BV) ~ log(B0) + (AS/AS.ref) * log(B1)

# Note. It is straightforward to show that M2 amd M1 are equivalent models, just
expressed
# in a different ways:
```

```
#     [M1] DV.BV ~ B0 * exp(B1'*AS) = B0 * exp((ref.AS*B1')*(AS/AS.ref)) =
#                B0 + B1^(AS/AS.ref),
# where B1 = exp(ref.AS*B1'). The parameters in model [M2] are simpler to use.
#

# For example, take the Tang and Vashishth (2007) plot. At AS = 0 and ref.AS = 2,
# DV/BV values are about 0.02 and 0.22, so the model is roughly
#           DV.BV ~ 0.02 * (0.22/0.02)*(AS/2) = 0.02 * 11^(AS/2),
# saying that DV/BV is 0.02 at 0% Apparent Strain, 0.02*11^(0.50) = 0.02*sqrt(11) =
# 0.066 at AS = 1%, and 0.02*11 = 0.22 at AS = 2%.
#
# Note. This is nearly identical to the fit shown in T&V's Figure 4:
#      DV.BV = 0.02 * exp(1.2*AS) = 0.02 * exp(2.4)^(AS/2) = 0.02 * 11.02*(AS/2).
#
# Specifically, this function allows for varying the
#   (1) number and position of the AS design points. Note. Using two design points,
AS=0 vs AS=ref.AS,
# sets up a two-group problem, which is handled here by a simple t-test whether the
geometric mean of
# [DV.BV(AS=ref.AS)]/[DV.BV(AS=0)] = 1.0. See example.
#   (2) sample size at each design point.
#   (3) values for B0 and B1.
#   (3) amount of totalnoise, expressed as 95% relative spread
#
# All relevant parameter values are modifiable in the function call; see examples.


# Summary.
# ========
#     AS.pts: Apparent Strength design points, such as
#           AS.pts = c(0, 0.5, 0.75, 1.00, 1.25 1.50, 1.75, 2.00, 2.25)

#     ref.AS: reference design point, single value, such as ref.AS = 2.

#        n: number of independent donors used at the given design points.

#        B0: geometric mean of DV/BV when % Apparent Strain = 0, gmean(AS=0).
#
#        B1: gmean(AS=ref.AS)/gmean(AS=0), the fold change in DV/BV when
#             going from AS = 0 to some reference point for AS, ref.AS > 0.

# TotalNoise.RS95:
#             the 95% relative spread of the total randomness of each DV/BV
#             measurement.
#
```

```
#    ConfLevel: Sets level of confidence intervals (e.g. 0.95).
#
#      B1.null: sets null for testing B1. The usual value is 1.0. "Rejecting"
#               this says there is some relationship between AS and DV/BV. So what?


# Simulation results, per trial
# ============================

#   B1.est: estimate of fold change in DV/BV from AS = 0 to ref.AS

#    B1.lo: lower 95% confidence limit for B1.
#    B1.hi: upper 95% confidence limit for B1.
#
#  p.value: one-sided p.value for testing
#             Ho: B1 = B1.null
#             Ha: B1 > B1.null


DV.BVcurve = function(
  title="a case with no title",
  AS.pts,           # design points for Apparent Strain
  AS.ref,           # reference design point for Apparent Strain
  n,               # number of donors used at each design point
  B0,               # true geometric mean of DV/BV at AS = 0, gmean(0)
  B1,               # true gmean(ref.AS)/gmean(0)
  TotalNoise.RS95,   # the 95% RS of the total randomness of each DV/BV
measurement.
  ConfLevel=0.95,    # level of confidence interval
  B1.null=1,         # sets null for testing B1
  Print=TRUE,        # set to FALSE to suppress printing
  Plot=TRUE,         # set to FALSE to suppress plotting of Trial 1 data
  LogScaleDV.BV=F,   # set to TRUE to log scale DV/BV axis
  NewWindow=TRUE,    # set to FALSE to re-use the same plot window
  seed=1234955,      # a seed to start simulations
  Ntrials=1000)      # number of trials in simulations
 {

 if (Print) {
   cat("\n\n",title,"\n")
   cat("\nDesign Points:", AS.pts)
   cat("\nRefefence Point:", AS.ref, "\n")
   cat("\nn:", n, "\n\n")
   print(data.frame(B0, B1, TotalNoise.RS95),row.names="")
   cat("\n")
```

```
  print(data.frame(ConfLevel,B1.null,row.names="")); cat("\n")
  print(data.frame(Print,seed,Ntrials,row.names="")); cat("\n")
  print(data.frame(Plot,LogScaleDV.BV,NewWindow,row.names="")); cat("\n")
}

set.seed(seed)
logB1.est <- p.value <- RelSpanCI <- rep(NA,Ntrials)
logB1.CLims <- matrix(NA, nrow=Ntrials, ncol=2)
K <- length(AS.pts)
Ntotal <- sum(n)

if (min(c(B0, B1, TotalNoise.RS95)) <= 0) {
  stop ("B0, B1, TotalNoise.RS95 must be positive.")
  }

if (TotalNoise.RS95 <= 1) {
                        stop ("TotalNoise.RS95 must exceed 1.0.")
}

SD.logTNoise <- log(TotalNoise.RS95)/(1.96*2)

DV.BV <- rep(NA, Ntotal)
AS <- rep(AS.pts, n)
X.AS <- AS/AS.ref
logB0 <- log(B0)
logB1 <- log(B1)

logB1.est <- logB1.CIlo <- logB1.CIhi <- B1.pvalue <- rep(NA, Ntrials)

logB1.null <- log(B1.null)
dfE <- Ntotal - 2

for (iTrial in 1:Ntrials)  {
  logDV.BV <- logB0 + X.AS*logB1 + rnorm(Ntotal, 0, SD.logTNoise)
  DV.BV <- exp(logDV.BV)

  fit <- lm(logDV.BV ~ X.AS)
  logB1.est[iTrial] <- fit$coefficients[2]
  logB1.CLims[iTrial,] <- confint(fit, level=ConfLevel)[2,]
  SE.B1 <- summary(fit)$coefficients[2,2]
  t.stat <- (fit$coefficients[2] - logB1.null)/SE.B1
  B1.pvalue[iTrial] <- pt(abs(t.stat), dfE, lower.tail = FALSE)  # one-sided

  if (Print & (iTrial==1)) {
     cat("\n\nFirst specimen in first trial\n")
```

```
    print (data.frame(AS, X.AS, DV.BV=round(exp(logDV.BV),3)))
    cat("\n")
    print(summary(fit))
  }

  if (Plot & (iTrial==1)) {  # plot Trial 1 data
        X <- seq(0,100)/(100/max(AS))
        Y <- exp(fit$coefficients[1] + (X/AS.ref)*fit$coefficients[2])
        plottitle <- "Data and Fit for Trial #1"
        if (NewWindow) {
    try(windows(width=6, height=4), silent=T)
    try(quartz(width=6, height=4), silent=T)
   }

        Ymin <- 0
        LogScaling = ""
        Ylabel <- "Damaged Volume Fraction (DV/BV)"
        Xmax <- max(AS)
        Ymax <- max(DV.BV)

        if (LogScaleDV.BV) {
                LogScaling = "y"
                DV.BV[DV.BV < 0.001] = 0.001
                Ymin <- min(0.001, min(DV.BV))
                Ylabel <- "Damaged Volume Fraction (DV/BV; log-scaled)"
                }

   plot(DV.BV ~ AS,
                ylim= c(Ymin, Ymax),
         xlim= c(0, Xmax),
                ylab=Ylabel,
                xlab = "Apparent Strain (%)",
                pch = "o",
                xaxt="n",
                log=LogScaling,
                las=1,
                cex.axis= 0.75, cex.lab=0.85,
                main=plottitle)
      axis(1, at=c(0, 0.5, 1, 1.5, 2, 2.5), cex.axis=0.75)
      lines(X,Y)
      }
  } # end for (iTrial in 1:Ntrials)

B1.pvalue.char <- as.character(round(B1.pvalue, 4))
B1.pvalue.char[B1.pvalue < 0.0001] <- "<0.0001"
```

```
  if (Print) {
  cat("\n\nResults for first 10 trials\n")

  first10 <- data.frame(B1.est=round(exp(logB1.est),3),
                    B1.CLimLo=round(exp(logB1.CLims[,1]),3),

          B1.CLimHi=round(exp(logB1.CLims[,2]),3),

          pvalue=B1.pvalue.char,
                    CIspan=round(exp(logB1.CLims[,2] -
                              logB1.CLims[,1]),3))[1:10,]
  print(first10)
  }

# Preliminary code to analyze the simulation study
{
  cat("\n\nSummary of B1 estimates; true =", B1, "\n")
  print(round(exp(summary(logB1.est)),2))
  cat("\n\nSummary of B1 lower confidence limits\n")
  print(round(exp(summary(logB1.CLims[,1])),2))
  cat("\n\nSummary of B1 upper confidence limits\n")
  print(round(exp(summary(logB1.CLims[,2])),2))
  cat("\n\nSummary of CI span = B1.CIhi/B1.CIlo\n")
  print(round(exp(summary(logB1.CLims[,2] -
          logB1.CLims[,1])),2))
  cat("\n\nSummary of B1 p.values; one-sided test of B1 > ", B1.null, "\n")
  print(summary(B1.pvalue,3))
}

} # end DV.BVcurve ()
```

WORKS CITED

[1] Burr D. Microdamage and bone strength. Osteoporos Int 2003;14 Suppl 5: S67-72.
[2] Schaffler MB. Role of bone turnover in microdamage. Osteoporos Int 2003;14 Suppl 5: S73-7; discussion S77-80.
[3] Tang SY, Vashishth D. A non-invasive in vitro technique for the three-dimensional quantification of microdamage in trabecular bone. Bone 2007;40: 1259-64.
[4] Wang X, Masse DB, Leng H, Hess KP, Ross RD, Roeder RK, Niebur GL. Detection of trabecular bone microdamage by micro-computed tomography. J Biomech 2007;40: 3397-403.
[5] Bigley RF, Singh M, Hernandez CJ, Kazakia GJ, Martin RB, Keaveny TM. Validity of serial milling-based imaging system for microdamage quantification. Bone 2008;42: 212-5.
[6] Slyfield CR, Tkachenko EV, Fischer SE, Ehlert KM, Yi IH, Jekir MG, O'Brien RG, Keaveny TM, Hernandez CJ. Mechanical failure begins preferentially near resorption cavities in human vertebral cancellous bone under compression. Bone 2012;50: 1281-7.
[7] Hernandez CJ, Ramsey DS, Dux SJ, Chu EH, Rimnac CM. Irradiation does not modify mechanical properties of cancellous bone under compression. Clin Orthop Relat Res 2012;470: 2488-95.
[8] Martin RB, Yeh OC, Fyhrie DP. On sampling bones for microcracks. Bone 2007;40: 1159-65.
[9] Glantz SA, Slinker BK. Primer of Applied Regression & Analysis of Variance. 2nd ed. New York: McGraw-Hill; 2001.
[10] Tang SY, Vashishth D. Non-enzymatic glycation alters microdamage formation in human cancellous bone. Bone 2010;46: 148-54.
[11] Karim L, Vashishth D. Role of trabecular microarchitecture in the formation, accumulation, and morphology of microdamage in human cancellous bone. J Orthop Res 2011;29: 1739-44.
[12] Keaveny TM, Wachtel EF, Kopperdahl DL. Mechanical behavior of human trabecular bone after overloading. J Orthop Res 1999;17: 346-53.
[13] Connor JT. The value of a p-valueless paper. Am J Gastroenterol 2004;99: 1638-40.
[14] Poole C. Low P-values or narrow confidence intervals: which are more durable? Epidemiology 2001;12: 291-4.
[15] Dmitrienko A, Chuang-Stein C, D'Agostino RB. Pharmaceutical statistics using SAS : a practical guide. Cary, N.C.: SAS Institute; 2007.
[16] O'Neal JM, Diab T, Allen MR, Vidakovic B, Burr DB, Guldberg RE. One year of alendronate treatment lowers microstructural stresses associated with trabecular microdamage initiation. Bone 2010;47: 241-7.
[17] Dux SJ, Ramsey D, Chu EH, Rimnac CM, Hernandez CJ. Alterations in damage processes in dense cancellous bone following gamma-radiation sterilization. J Biomech 2010;43: 1509-13.

[18] Nagaraja S, Lin AS, Guldberg RE. Age-related changes in trabecular bone microdamage initiation. Bone 2007;40: 973-80.
[19] Nagaraja S, Couse TL, Guldberg RE. Trabecular bone microdamage and microstructural stresses under uniaxial compression. J Biomech 2005;38: 707-16.
[20] Wang X, Niebur GL. Microdamage propagation in trabecular bone due to changes in loading mode. J Biomech 2006;39: 781-90.
[21] Wang X, Guyette J, Liu X, Roeder RK, Niebur GL. Axial-shear interaction effects on microdamage in bovine tibial trabecular bone. Eur J Morphol 2005;42: 61-70.
[22] Moore TL, Gibson LJ. Fatigue microdamage in bovine trabecular bone. J Biomech Eng 2003;125: 769-76.
[23] Moore TLA, Gibson LJ. Microdamage accumulation in bovine trabecular bone in uniaxial compression. J Biomech Eng 2002;124: 63-71.